

B6

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
31 July 2003 (31.07.2003)

PCT

(10) International Publication Number  
WO 03/062396 A2

(51) International Patent Classification<sup>7</sup>: C12N  
(21) International Application Number: PCT/US03/02136  
(22) International Filing Date: 22 January 2003 (22.01.2003)  
(25) Filing Language: English  
(26) Publication Language: English  
(30) Priority Data:  
60/351,570 23 January 2002 (23.01.2002) US  
(71) Applicant (*for all designated States except US*): ARENA  
PHARMACEUTICALS, INC. [US/US]; 6166 Nancy  
Ridge Drive, San Diego, CA 92121 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE,  
SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,  
VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,  
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI,  
SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN,  
GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventor; and  
(75) Inventor/Applicant (*for US only*): LIAW, Chen [US/US];  
5409 Harvest Run Drive, San Diego, CA 92130 (US).  
(74) Agents: BRADFUTE, David et al.; Arena Pharmaceuti-  
cals, Inc., 6166 Nancy Ridge Drive, San Diego, CA 92121  
(US).

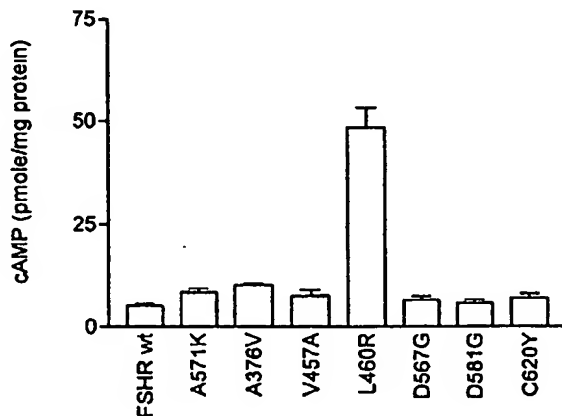
Published:

— without international search report and to be republished  
upon receipt of that report

*For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*

(54) Title: NON-ENDOGENOUS VERSIONS OF HUMAN G PROTEIN-COUPLED RECEPTOR: FSHR

FSHR Cyclase Assay  
In 293 Cells



(57) Abstract: The invention disclosed in this patent document relates to transmembrane receptors, particularly to a human G protein-coupled receptor, more particularly to a follicle stimulating hormone receptor (FSHR), and most particularly to mutated (non-endogenous) versions of the human FSHR for evidence of constitutive activity.

WO 03/062396 A2

**NON-ENDOGENOUS VERSIONS OF  
HUMAN G PROTEIN-COUPLED RECEPTOR: FSHR**

**CROSS-REFERENCE TO RELATED APPLICATIONS**

This application claims priority benefit of U.S. Provisional Application Number 06/351,570,  
5 filed January 23, 2002, which is hereby incorporated by reference in its entirety.

**FIELD OF THE INVENTION**

The invention disclosed in this patent document relates to transmembrane receptors, and more particularly to a G protein-coupled receptor ("GPCR") for which the endogenous ligand has  
10 been identified; and specifically to a follicle stimulating hormone receptor ("FSHR") that has been altered to establish constitutive activity of the receptor. In some embodiments the altered versions of FSHR are used, *inter alia*, for the direct identification of candidate compounds such as receptor agonists, inverse agonists, partial agonist or antagonist for use in, for example and not limitation, ovulation, osteoporosis, menopausal women, prostate cancer, and Polycystic Ovary Syndrome  
15 (PCOS) which can ultimately lead to non-insulin dependent diabetes (NIDDM). Candidate compounds identified according to the methods disclosed herein may be useful in primates, including but not limited to, humans and non-human primates; as well other mammals, including but not limited to, dogs and cats, rats, mice, horses, sheep, pigs, cows, and other mammals that are considered endangered.

20

**BACKGROUND OF THE INVENTION**

**A. G Protein-Coupled Receptors**

[0001] Although a number of receptor classes exist in humans, by far the most abundant  
25 and therapeutically relevant is represented by the G protein-coupled receptor (GPCR) class. It is estimated that there are some 30,000-40,000 genes within the human genome, and of these, approximately 2% are estimated to code for GPCRs. Receptors, including GPCRs, for which the endogenous ligand has been identified, are referred to as "known" receptors, while receptors for which the endogenous ligand has not been identified are referred to as "orphan" receptors.

[0002] GPCRs represent an important area for the development of pharmaceutical products: from approximately 20 of the 100 known GPCRs, approximately 60% of all prescription pharmaceuticals have been developed. For example, in 1999, of the top 100 brand name prescription drugs, the following drugs interact with GPCRs (the primary diseases and/or disorders treated related to the drug is indicated in parentheses):

Claritin® (allergies)	Prozac® (depression)	Vasotec® (hypertension)
Paxil® (depression)	Zoloft® (depression)	Zyprexa® (psychotic disorder)
Cozaar® (hypertension)	Imitrex® (migraine)	Zantac® (reflux)
Propulsid® (reflux disease)	Risperdal® (schizophrenia)	Serevent® (asthma)
10 Pepcid® (reflux)	Gaster® (ulcers)	Atrovent® (bronchospasm)
Effexor® (depression)	Depakote® (epilepsy)	Cardura® (prostatic hypertrophy)
Allegra® (allergies)	Lupron® (prostate cancer)	Zoladex® (prostate cancer)
Diprivan® (anesthesia)	BuSpar® (anxiety)	Ventolin® (bronchospasm)
Hytrin® (hypertension)	Wellbutrin® (depression)	Zyrtec® (rhinitis)
15 Plavix® (MI/stroke)	Toprol-XL® (hypertension)	Tenormin® (angina)
Xalatan® (glaucoma)	Singulair® (asthma)	Diovan® (hypertension)
Harnal® (prostatic hyperplasia)		
(Med Ad News 1999 Data).		

[0003] GPCRs share a common structural motif, having seven sequences of between 22 to 24 hydrophobic amino acids that form seven alpha helices, each of which spans the membrane (each span is identified by number, *i.e.*, transmembrane-1 (TM-1), transmembrane-2 (TM-2), *etc.*). The transmembrane helices are joined by strands of amino acids between transmembrane-2 and transmembrane-3, transmembrane-4 and transmembrane-5, and transmembrane-6 and transmembrane-7 on the exterior, or "extracellular" side, of the cell membrane (these are referred to as "extracellular" regions 1, 2 and 3 (EC-1, EC-2 and EC-3), respectively). The transmembrane helices are also joined by strands of amino acids between transmembrane-1 and transmembrane-2, transmembrane-3 and transmembrane-4, and transmembrane-5 and transmembrane-6 on the interior, or "intracellular" side, of the cell membrane (these are referred to as "intracellular" regions 1, 2 and 3 (IC-1, IC-2 and IC-3), respectively). The "carboxy" ("C")

terminus of the receptor lies in the intracellular space within the cell, and the "amino" ("N") terminus of the receptor lies in the extracellular space outside of the cell.

[0004] Generally, when an endogenous ligand binds with the receptor (often referred to as "activation" of the receptor), there is a change in the conformation of the intracellular region that allows for coupling between the intracellular region and an intracellular "G-protein." It has been reported that GPCRs are "promiscuous" with respect to G proteins, *i.e.*, that a GPCR can interact with more than one G protein. *See*, Kenakin, T., 43 *Life Sciences* 1095 (1988). Although other G proteins exist, currently, G<sub>s</sub>, G<sub>i</sub>, G<sub>o</sub>, G<sub>q</sub>, G<sub>12</sub> and G<sub>13</sub> are G proteins that have been identified. Ligand-activated GPCR coupling with the G-protein initiates a signaling cascade process (referred to as "signal transduction"). Under normal conditions, signal transduction ultimately results in cellular activation or cellular inhibition. Although not wishing to be bound to theory, it is thought that the IC-3 loop as well as the carboxy terminus of the receptor interact with the G protein.

[0005] Under physiological conditions, GPCRs exist in the cell membrane in equilibrium between two different conformations: an "inactive" state and an "active" state. A receptor in an inactive state is unable to link to the intracellular signaling transduction pathway to initiate signal transduction leading to a biological response. Changing the receptor conformation to the active state allows linkage to the transduction pathway (via the G-protein) and produces a biological response.

[0006] A receptor may be stabilized in an active state by a ligand or a compound such as a drug. Recent discoveries, including but not exclusively limited to modifications to the amino acid sequence of the receptor, provide means other than ligands or drugs to promote and stabilize the receptor in the active state conformation. These means effectively stabilize the receptor in an active state by simulating the effect of a ligand binding to the receptor. Stabilization by such ligand-independent means is termed "constitutive receptor activation."

#### B. Follicle Stimulating Hormone Receptor ("FSHR")

[0007] The follicle stimulating hormone receptor (FSHR) is known to be a G protein-coupled receptor whereby the natural ligand has been identified as the follicle stimulating hormone (FSH), a heterodimeric glycoprotein hormone. FSH shares structural similarities with

the leutinizing hormone (LH) and the thyroid stimulating hormone (TSH), both of which are produced in the pituitary gland. .

5 [0008] FSHR has been determined to be expressed in the testicular Sertoli cells and ovarian granulose cells. Similarly, LHR has been determined to be expressed in the Leydig cells in the testis, the theca cells in the ovary, the granulosa cells, the corpus luteum cells and the interstitial cells, and has been reported to play a role in reproductive physiology. When activated these receptors stimulate an increase in the activity of adenylyl cyclase, which in turn causes increased steroid synthesis and secretion.

10

#### SUMMARY OF THE INVENTION

[0009] The present invention discloses nucleic acid molecules and the proteins for a non-endogenous, constitutively activated versions of human FSH receptor, referred to herein as, A376V, V457A, L460R, D567G, A571K, D581G, and C620Y. The L460R receptor has been determined to be a constitutively active form of the human FSHR created by a point mutation  
15 from a leucine amino acid residue located at position 460 to an arginine residue.

[0010] The present invention relates to non-endogenous, constitutively activated versions of the human follicle stimulating hormone receptor ("FSHR") and various uses of such receptor. In some embodiments, FSHR has an amino acid sequence of SEQ ID NO: 26. In some embodiments, FSHR is encoded by a nucleotide sequence of SEQ ID NO: 25.

20 [0011] In further aspects the present invention is directed to plasmids comprising a vector and a cDNA having SEQ.ID.NO: 25.

[0012] In some aspects the present invention is directed to host cells comprising a plasmid wherein the plasmid comprises a vector and a cDNA having SEQ.ID.NO: 25.

25 [0013] In additional aspects the present invention is directed to methods for directly identifying a non-endogenous candidate compound as an agonist, an inverse agonist, partial agonist or an antagonist to an endogenous FSHR. The methods comprise the steps of: (a) subjecting the endogenous FSHR to constitutive receptor activation to create a non-endogenous, constitutively activated FSHR; (b) contacting the non-endogenous candidate compound with the non-endogenous, constitutively activated FSHR; and (c) identifying the non-

endogenous candidate compound as an agonist, an inverse agonist, a partial agonist or an antagonist to the constitutively activated FSHR by measuring at least a 20% difference in an intracellular signal induced by the contacted compound as compared with an intracellular signal in the absence of the contacted compound. These identified candidate compounds can then be  
5 utilized in pharmaceutical composition(s) for treatment of disease and disorders which are related to the human FSH receptor.

[0014] In additional aspects the present invention is directed to compounds identified by the methods set forth above and described below.

[0015] In additional aspects the present invention is directed to compositions, including  
10 pharmaceutical compositions, comprising compounds directly identified by the methods of the present invention.

[0016] In some aspects the present invention is directed to methods of modulating a physiological process comprising subjecting an endogenous FSHR to constitutive receptor activation to create a non-endogenous, constitutively activated FSHR. The physiological process  
15 is thereby modulated. In some embodiments, the endogenous FSHR has an amino acid sequence of SEQ ID NO: 26.

[0017] In some embodiments the physiological process is selected from the group consisting of ovulation, osteoporosis, menopausal women, prostate cancer, and Polycystic Ovary Syndrome (PCOS) which can ultimately lead to non-insulin dependent diabetes (NIDDM).

[0018] In additional aspects, the present invention is directed to methods of modulating a physiological process comprising: (a) subjecting an endogenous FSHR to constitutive receptor activation to create a non-endogenous constitutively activated FSHR; and (b) contacting the non-endogenous, constitutively activated FSHR with a non-endogenous agonist, inverse agonist, partial agonist or antagonist of said FSHR. The physiological process is thereby modulated. In  
20 some embodiments, the endogenous FSHR has an amino acid sequence of SEQ ID NO: 26. In some embodiments the physiological process is selected from the group consisting of ovulation, osteoporosis, menopausal women, prostate cancer, and Polycystic Ovary Syndrome (PCOS) which can ultimately lead to non-insulin dependent diabetes (NIDDM).

[0019] In some aspects the present invention is directed to methods for directly  
30 identifying a non-endogenous candidate compound as a compound having activity selected from

the group consisting of inverse agonist activity and agonist activity, to an endogenous, constitutively active G protein coupled cell surface receptor (GPCR) comprising the steps of: (a) contacting a non-endogenous candidate compound with a GPCR Fusion Protein, the GPCR Fusion Protein comprising the endogenous, constitutively active FSHR and a G protein; and  
5 (b) identifying the non-endogenous candidate compound as an agonist, an inverse agonist, partial agonist or antagonist to the endogenous constitutively activated FSHR by measuring at least a 20% difference in an intracellular signal induced by the contacted compound as compared with an intracellular signal in the absence of the contacted compound.

[0020] In additional aspects the present invention is directed to methods for directly  
10 identifying a non-endogenous candidate compound as a compound having activity selected from the group consisting of inverse agonist activity and agonist activity, to an endogenous, constitutively active G protein coupled cell surface receptor (GPCR) comprising the steps of: (a) contacting a non-endogenous candidate compound with a GPCR Fusion Protein, the GPCR Fusion Protein comprising the endogenous, constitutively active FSHR and a G protein; and  
15 (b) determining whether a receptor functionality is modulated, wherein a change in receptor functionality is indicative of the candidate compound being an agonist, inverse agonist, partial agonist or antagonist of said endogenous, constitutively active FSHR.

[0021] In some aspects the present invention is directed to GPCR Fusion Protein constructs comprising a constitutively active G protein coupled receptor and a G protein. In some  
20 embodiments, the constitutively active G protein coupled receptor is non-endogenous. In some embodiments, the GPCR Fusion Protein construct comprises constitutively active G protein coupled receptor comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 26. In some embodiments, the said G protein is Gs.

[0022] In some aspects the present invention is directed to methods for modulating a  
25 physiological process in primates, including but not limited to humans and non-human primates; as well as other mammals, including but not limited to, dogs, cats, rats, mice, horses, sheep, pigs, cows, and other mammals that are considered to be endangered. The methods comprise the steps of: (a) subjecting an endogenous FSHR to constitutive receptor activation to create a non-endogenous, constitutively activated FSHR; (b) contacting the non-endogenous candidate  
30 compound with the non-endogenous, constitutively activated GPCR; (c) identifying the non-endogenous candidate compound as an agonist, an inverse agonist, a partial agonist or antagonist to the non-endogenous constitutively activated FSHR by measuring at least a 20% difference in

an intracellular signal induced by the contacted compound as compared with an intracellular signal in the absence of the contacted compound; and (d) contacting the mammal with the inverse agonist or agonist; whereby the physiological process is modulated.

[0023] In other aspects the present invention is directed to a mammal comprising a non-endogenous, constitutively activated G protein-coupled receptor (GPCR). In some embodiments, the G protein-coupled receptor has an amino acid sequence of SEQ ID NO: 26. In some embodiments, the G protein-coupled receptor is encoded by a nucleotide sequence of SEQ ID NO: 25.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0024] In the following figures, bold typeface indicates the location of the mutation in the non-endogenous, constitutively activated receptor relative to the corresponding endogenous receptor.

[0025] Figure 1 depicts the results of a second messenger cell-based cyclic AMP assay providing comparative results for constitutive signaling of endogenous FSHR ("FSHRwt"), non-endogenous versions of FSHR ("L460R", "A376V", "V457A", "L460R", "D567G", "A571K", "D581G", and "C620Y") and a control vector ("CMV").

[0026] Figure 2 depicts the results of a second messenger cAMP accumulation in an Alpha Screen comparing the results of endogenous FSHR ("WT"), non-endogenous versions of FSHR ("L460R") and a control vector ("CMV"). The data evidences that the L460R version of the FSHR is constitutively activated.

[0027] Figure 3 depicts the results of cAMP accumulation in an Alpha Screen analysis of the endogenous FSHR ("WT") compared with the non-endogenous FSHR ("L460R") and a control vector ("CMV") in the presence of Compound A. Compound A binds to the WT receptor at an EC50 of about 3nM, while Compound A binds the L460R version of FSHR at about 7μM. This data evidences that Compound A has a better efficacy for the non-endogenous, constitutively activated version of FSHR (L460) than the WT receptor.

## DETAILED DESCRIPTION

The scientific literature that has evolved around receptors has adopted a number of terms to refer to ligands having various effects on receptors. For clarity and consistency, the following definitions will be used throughout this patent document. To the extent that these definitions conflict with other definitions for these terms, the following definitions shall control:

- 5 [0028] **AGONISTS** shall mean materials (*e.g.*, ligands, candidate compounds) that activate the intracellular response when they bind to the receptor, or enhance GTP binding to membranes. In some embodiments, **AGONISTS** are those materials not previously known to activate the intracellular response when they bind to the receptor or to enhance GTP binding to membranes.
- 10 [0029] **ALLOSTERIC MODULATORS** shall mean materials (*e.g.*, ligands, candidate compounds) that affect the functional activity of the receptor but which do not inhibit the endogenous ligand from binding to the receptor. Allosteric modulators include inverse agonists, partial agonists and agonists.

[0030] **AMINO ACID ABBREVIATIONS** used herein are set out in Table A:

15

TABLE A

ALANINE	ALA	A
ARGININE	ARG	R
ASPARAGINE	ASN	N
ASPARTIC ACID	ASP	D
CYSTEINE	CYS	C
GLUTAMIC ACID	GLU	E
GLUTAMINE	GLN	Q
GLYCINE	GLY	G
HISTIDINE	HIS	H
ISOLEUCINE	ILE	I
LEUCINE	LEU	L
LYSINE	LYS	K
METHIONINE	MET	M

PHENYLALANINE	PHE	F
PROLINE	PRO	P
SERINE	SER	S
THREONINE	THR	T
TRYPTOPHAN	TRP	W
TYROSINE	TYR	Y
VALINE	VAL	V

[0031] **ANTAGONIST** shall mean materials (e.g., ligands, candidate compounds) that competitively bind to the receptor at the same site as the agonists but which do not activate the intracellular response initiated by the active form of the receptor, and can thereby inhibit the intracellular responses by agonists. **ANTAGONISTS** do not diminish the baseline intracellular response in the absence of an agonist. In some embodiments, **ANTAGONISTS** are those materials not previously known to activate the intracellular response when they bind to the receptor or to enhance GTP binding to membranes.

[0032] **CANDIDATE COMPOUND** shall mean a molecule (for example, and not limitation, a chemical compound) that is amenable to a screening technique. Preferably, the phrase "candidate compound" does not include compounds which were publicly known to be compounds selected from the group consisting of inverse agonist, agonist or antagonist to a receptor, as previously determined by an indirect identification process ("indirectly identified compound"); more preferably, not including an indirectly identified compound which has previously been determined to have therapeutic efficacy in at least one mammal; and, most preferably, not including an indirectly identified compound which has previously been determined to have therapeutic utility in humans.

[0033] **COMPOSITION** means a material comprising at least one component; a "pharmaceutical composition" is an example of a composition.

[0034] **COMPOUND EFFICACY** shall mean a measurement of the ability of a compound to inhibit or stimulate receptor functionality; i.e. the ability to activate/inhibit a signal transduction pathway, as opposed to receptor binding affinity. Exemplary means of detecting compound efficacy are disclosed in the Example section of this patent document.

[0035] **CODON** shall mean a grouping of three nucleotides (or equivalents to nucleotides) which generally comprise a nucleoside (adenosine (A), guanosine (G), cytidine (C), uridine (U) and thymidine (T)) coupled to a phosphate group and which, when translated, encodes an amino acid.

5 [0036] **CONSTITUTIVELY ACTIVATED RECEPTOR** shall mean a receptor subjected to constitutive receptor activation. A constitutively activated receptor can be endogenous or non-endogenous.

[0037] **CONSTITUTIVE RECEPTOR ACTIVATION** shall mean stabilization of a receptor in the active state by means other than binding of the receptor with its ligand or a  
10 chemical equivalent thereof.

[0038] **CONTACT** or **CONTACTING** shall mean bringing at least two moieties together, whether in an in vitro system or an in vivo system.

[0039] **DECREASE** is used to refer to a reduction in a measurable quantity and is used synonymously with the terms "reduce", "diminish", "lower", and "lessen".

15 [0040] **DIRECTLY IDENTIFYING** or **DIRECTLY IDENTIFIED**, in relationship to the phrase "candidate compound", shall mean the screening of a candidate compound against a constitutively activated receptor, preferably a constitutively activated orphan receptor, and most preferably against a constitutively activated G protein-coupled cell surface orphan receptor, and assessing the compound efficacy of such compound. This phrase is, under no circumstances, to be  
20 interpreted or understood to be encompassed by or to encompass the phrase "indirectly identifying" or "indirectly identified."

[0041] **ENDOGENOUS** shall mean a material that a mammal naturally produces. **ENDOGENOUS** in reference to, for example and not limitation, the term "receptor," shall mean that which is naturally produced by a mammal (for example, and not limitation, a human) or a  
25 virus. By contrast, the term **NON-ENDOGENOUS** in this context shall mean that which is not naturally produced by a mammal (for example, and not limitation, a human) or a virus. For example, and not limitation, a receptor which is not constitutively active in its endogenous form, but when manipulated becomes constitutively active, is most preferably referred to herein as a "non-endogenous, constitutively activated receptor." Both terms can be utilized to describe both  
30 "in vivo" and "in vitro" systems. For example, and not limitation, in a screening approach, the

endogenous or non-endogenous receptor may be in reference to an in vitro screening system. As a further example and not limitation, where the genome of a mammal has been manipulated to include a non-endogenous constitutively activated receptor, screening of a candidate compound by means of an in vivo system is viable.

5    **[0042]**       **EXPRESSION VECTOR** shall refer to the molecules that comprise a nucleic acid sequence which encode one or more desired polypeptides and which include initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression.

10   **[0043]**       **G PROTEIN COUPLED RECEPTOR FUSION PROTEIN** and **GPCR FUSION PROTEIN**, in the context of the invention disclosed herein, each mean a non-endogenous protein comprising an endogenous, constitutively activate GPCR or a non-endogenous, constitutively activated GPCR fused to at least one G protein, most preferably the alpha ( $\alpha$ ) subunit of such G protein (this being the subunit that binds GTP), with the G protein preferably being of the same type as the G protein that naturally couples with endogenous orphan  
15   GPCR. For example, and not limitation, in an endogenous state, if the G protein "G $\alpha$ " is the predominate G protein that couples with the GPCR, a GPCR Fusion Protein based upon the specific GPCR would be a non-endogenous protein comprising the GPCR fused to G $\alpha$ ; in some circumstances, as will be set forth below, a non-predominant G protein can be fused to the GPCR. The G protein can be fused directly to the C-terminus of the constitutively active GPCR or there  
20   may be spacers between the two.

25   **[0044]**       **HOST CELL** shall mean a cell capable of having a Plasmid and/or Vector incorporated therein. In the case of a prokaryotic Host Cell, a Plasmid is typically replicated as a autonomous molecule as the Host Cell replicates (generally, the Plasmid is thereafter isolated for introduction into a eukaryotic Host Cell); in the case of a eukaryotic Host Cell, a Plasmid is integrated into the cellular DNA of the Host Cell such that when the eukaryotic Host Cell replicates, the Plasmid replicates. In some embodiments the Host Cell is eukaryotic, more preferably, mammalian, and most preferably selected from the group consisting of 293, 293T and COS-7 cells.

30   **[0045]**       **INDIRECTLY IDENTIFYING** or **INDIRECTLY IDENTIFIED** means the traditional approach to the drug discovery process involving identification of an endogenous ligand specific for an endogenous receptor, screening of candidate compounds against the

receptor for determination of those which interfere and/or compete with the ligand-receptor interaction, and assessing the efficacy of the compound for affecting at least one second messenger pathway associated with the activated receptor.

5 [0046] **INHIBIT** or **INHIBITING**, in relationship to the term "response" shall mean that a response is decreased or prevented in the presence of a compound as opposed to in the absence of the compound.

10 [0047] **INVERSE AGONISTS** shall mean materials (*e.g.*, ligand, candidate compound) which bind to either the endogenous form of the receptor or to the constitutively activated form of the receptor, and which inhibit the baseline intracellular response initiated by the active form of the receptor below the normal base level of activity which is observed in the absence of agonists, or decrease GTP binding to membranes. Preferably, the baseline intracellular response is inhibited in the presence of the inverse agonist by at least 30%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, and most preferably at least 99% as compared with  
15 the baseline response in the absence of the inverse agonist.

[0048] **INTRACELLULAR SIGNAL** shall mean a detectable signal transduced by a receptor. Examples of intracellular signals are well-known to the art-skilled. Intracellular signals may be endogenous, *e.g.* an endogenous intracellular signal including without limitation second messengers; or non-endogenous, *e.g.* a non-endogenous intracellular signal including without  
20 limitation a engineered signal, *i.e.*,  $\beta$ -galactosidase, GUS, luciferase. Assays for detecting intracellular signals are known to those skilled in the art and include GTP $\gamma$ S assays, cAMP assays; CREB assays;  $\beta$ -galactosidase assays; luciferase assays; DAG assays; AP1 assays; IP<sub>3</sub> assays; and adenylyl cyclase assays. In some embodiments the term **INTRACELLULAR SIGNAL** is used synonymously with "reporter signal".

25 [0049] **KNOWN RECEPTOR** shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has been identified.

[0050] **LIGAND** shall mean a molecule specific for a naturally occurring receptor.

[0051] As used herein, the terms **MODULATE** or **MODIFY** are meant to refer to an increase or decrease in the amount, quality, or effect of a particular activity, function or molecule.

[0052] **MODULATE** shall mean an increase or decrease in an amount, quality, or effect of a particular activity or protein.

[0053] **MUTANT** or **MUTATION** in reference to an endogenous receptor's nucleic acid and/or amino acid sequence shall mean a specified change or changes to such endogenous sequences such that a mutated form of an endogenous non-constitutively activated receptor evidences constitutive activation of the receptor. In terms of equivalents to specific sequences, a subsequent mutated form of a human receptor is considered to be equivalent to a first mutation of the human receptor if (a) the level of constitutive activation of the subsequent mutated form of a human receptor is substantially the same as that evidenced by the first mutation of the receptor; and (b) the percent sequence (amino acid and/or nucleic acid) homology between the subsequent mutated form of the receptor and the first mutation of the receptor is at least 80%, at least 85%, at least 90%, at least 92%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, and most preferably at least 99%. In some embodiments, owing to the fact that some preferred cassettes disclosed herein for achieving constitutive activation include a single amino acid and/or codon change between the endogenous and the non-endogenous forms of the GPCR, it is preferred that the percent sequence homology should be at least 98%.

[0054] **NON-ORPHAN RECEPTOR** shall mean an endogenous naturally occurring molecule specific for an identified ligand wherein the binding of a ligand to a receptor activates an intracellular signaling pathway.

[0055] **ORPHAN RECEPTOR** shall mean an endogenous receptor for which the ligand specific for that receptor has not been identified or is not known.

[0056] **PARTIAL AGONISTS** shall mean materials (e.g., ligands, candidate compounds) that activate the intracellular response when they bind to the receptor to a lesser degree/extent than do agonists, or enhance GTP binding to membranes to a lesser degree/extent than do agonists. Preferably, the intracellular response is a lesser degree/extent than of an agonist by at least 95%, at least 80%, at least 70%, at least 60%, at least 65%, at least 50%, at least 45%, at least 40%, at least 38%, at least 35%, at least 34%, at least 33%, at least 32%, at least 31%, and most preferably at least 30% as compared with the baseline response of an agonist.

[0057] **PHARMACEUTICAL COMPOSITION** shall mean a composition comprising at least one active ingredient, whereby the composition is amenable to investigation for a specified, efficacious outcome in a mammal (for example, and not limitation, a human). Those of

ordinary skill in the art will understand and appreciate the techniques appropriate for determining whether an active ingredient has a desired efficacious outcome based upon the needs of the artisan.

[0058] **PLASMID** shall mean the combination of a Vector and cDNA. Generally, a  
5 Plasmid is introduced into a Host Cell for the purposes of replication and/or expression of the cDNA as a protein.

[0059] **RECEPTOR FUNCTIONALITY** shall refer to the normal operation of a receptor to receive a stimulus and moderate an effect in the cell, including, but not limited to regulating gene transcription, regulating the influx or efflux of ions, effecting a catalytic reaction,  
10 and/or modulating activity through G-proteins. **RECEPTOR FUNCTIONALITY** can readily be measured by the art skilled by measuring, without limitation, intracellular signals, ion influx or efflux, gene transcription, and effect of catalytic reaction.

[0060] **SECOND MESSENGER** shall mean an intracellular response produced as a result of receptor activation. A second messenger can include, for example, inositol triphosphate  
15 (IP<sub>3</sub>), diacylglycerol (DAG), cyclic AMP (cAMP), and cyclic GMP (cGMP). Second messenger response can be measured for a determination of receptor activation. In addition, second messenger response can be measured for the direct identification of candidate compounds, including for example, inverse agonists, partial agonists, agonists, and antagonists.

[0061] **SIGNAL TO NOISE RATIO** shall mean the signal generated in response to activation, amplification, or stimulation wherein the signal is above the background noise or the  
20 basal level in response to non-activation, non-amplification, or non-stimulation. In some preferred embodiments, the signal is at least 10%, preferably at least 20%, more preferably at least 30%, more preferably at least 40%, more preferably at least 50%, more preferably at least 60%, more preferably at least 70%, more preferably at least 80%, more preferably at least 90%,  
25 and most preferably at least 100% above background noise or basal level.

[0062] **SPACER** shall mean a translated number of amino acids that are located after the last codon or last amino acid of a gene, for example a GPCR of interest, but before the start codon or beginning regions of the G protein of interest, wherein the translated number amino acids are placed in-frame with the beginnings regions of the G protein of interest. The number of  
30 translated amino acids can be tailored according to the needs of the skilled artisan and is generally from about one amino acid, preferably two amino acids, more preferably three amino acids, more

preferably four amino acids, more preferably five amino acids, more preferably six amino acids, more preferably seven amino acids, more preferably eight amino acids, more preferably nine amino acids, more preferably ten amino acids, more preferably eleven amino acids, and even more preferably twelve amino acids.

- 5 [0063] **STIMULATE** or **STIMULATING**, in relationship to the term "response" shall mean that a response is increased in the presence of a compound as opposed to in the absence of the compound.

- [0064] **SUBJECTING AN ENDOGENOUS GPCR TO CONSTITUTIVE RECEPTOR ACTIVATION** shall refer to the steps through which a GPCR is constitutively  
10 activated.

[0065] **SUBJECT** shall mean primates, including but not limited to humans and non-human primates; as other mammals, including but not limited to, dogs, cats, rats, mice, horses, sheep, pigs, cows, and other mammals that are considered to be endangered.

- [0066] **SUBSTANTIALLY SIMILAR** shall refer to a result that is within 40% of a control result, preferably within 35%, more preferably within 30%, more preferably within 25%,  
15 more preferably within 20%, more preferably within 15%, more preferably within 10%, more preferably within 5%, more preferably within 2%, and most preferably within 1% of a control result. For example, in the context of receptor functionality, a test receptor may exhibit **SUBSTANTIALLY SIMILAR** results to a control receptor if the transduced signal, measured  
20 using a method taught herein or similar method known to the art-skilled, if within 40% of the signal produced by a control signal.

[0067] **VECTOR** in reference to cDNA shall mean a circular DNA capable of incorporating at least one cDNA and capable of incorporation into a Host Cell.

- [0068] The order of the following sections is set forth for presentational efficiency and is  
25 not intended, nor should be construed, as a limitation on the disclosure or the claims to follow.

## Introduction

[0069] Annually in the U.S. there are 2.4 million couples experiencing infertility that are potential candidates for treatment. Follicle stimulating hormone (FSH), either extracted from urine or produced by recombinant DNA technology, is a parenterally-administered protein

product used by specialists for induction ovulation and for controlled ovarian hyperstimulation (COH). Induction ovulation is necessary for in vitro fertilization process and treatment of PCOS; while COH is helpful in achieving healthier eggs, extra eggs and may increase the pregnancy rate for a woman.

- 5    [0070]       FSH and LH have been known to act on the ovary to stimulate steroid synthesis and secretion. FSH and LH are secreted by the pituitary and together play a central role in regulating the menstrual cycle and ovulation.

- 10   [0071]       In the normal menstrual cycle, there is a mid-cycle surge in LH concentration which is followed by ovulation. An elevated estrogen level, which is brought about by the endogenous secretion of LH and FSH, is required for the LH surge to occur. The estrogen mediates a positive feedback mechanism which results in the increased LH secretion. Oral contraceptive agents have been used by over 200 million women worldwide and by 1 of 4 women in the United States under the age of 45. Such agents are popular because of ease of administration, low pregnancy rate (less than 1 percent) and a relatively low incidence of side effects. Typically, oral contraceptives inhibit ovulation by suppressing FSH and LH secretion. As a consequence, the secretion of all ovarian steroids is also suppressed, including estrogen, progesterone and androgen. These agents also exert minor direct inhibitory effects on the reproductive tract, altering the cervical mucus, thereby decreasing sperm penetration and decreasing the motility and secretions of the fallopian tubes and uterus.

- 20   [0072]       The FSH receptor is expressed on testicular Sertoli cells and ovarian granulosa cells. While there has been a recognized need for providing essentially pure human FSH receptor, purification of naturally derived preparations is not practical and would likely be insufficient to permit determination of the amino acid sequence.

- 25   [0073]       Use of FSH is limited by its high cost, lack of oral dosing, and need of extensive monitoring by specialist physicians. Hence, identification of a non-peptidic small molecule substitute for FSH that could potentially be developed for oral administration is desirable.

- 30   [0074]       As described above, use of constitutively active forms of the G protein-coupled receptor FSHR, disclosed in the present patent document, can lead to an increase in steroid synthesis and secretion. Constitutively activated non-endogenous version of FSHR can be obtained, without limitation, by site-directed mutational methods. Constitutively active receptors useful for direct identification of candidate compounds are most preferably achieved by mutating

the receptor at a specific location, for example within transmembrane six (TM6) regions. Such mutations can produce a non-endogenous receptor that is constitutively activated, as evidenced by an increase in the functional activity of the receptor, for example, an increase in the level of second messenger activity.

5 [0075] As will be set forth and disclosed in greater detail below, utilization of several mutational approaches to modify the endogenous sequence of FSHR leads to constitutively activated versions of this receptor. These non-endogenous, constitutively activated version of FSHR can be utilized, *inter alia*, for the screening of candidate compounds to directly identify compounds which modulate processes and activities including, but not limited to, ovulation, 10 osteoporosis, menopausal women, prostate cancer, and Polycystic Ovary Syndrome (PCOS) which can ultimately lead to non-insulin dependent diabetes (NIDDM). Such physiological processes can further be modulated through, *inter alia*, subjecting an endogenous FSHR to constitutive receptor activation to create a non-endogenous, constitutively activated FSHR; and contacting the non-endogenous, constitutively activated FSHR with a non-endogenous agonist, 15 inverse agonist, partial agonist or antagonist of the receptor, or, in other embodiments, by subjecting an endogenous FSHR to constitutive receptor activation to create a non-endogenous, constitutively activated FSHR, whereby the physiological process is modulated.

#### B. Receptor Screening

[0076] Screening candidate compounds against a non-endogenous, constitutively 20 activated version of the GPCR disclosed herein allows for the direct identification of candidate compounds which act at the cell surface of the receptor, without requiring use of the receptor's endogenous ligand. This patent document discloses several mutational approaches for creating non-endogenous, constitutively activated versions of FSHR. With the disclosed techniques, one skilled in the art is credited with the ability to create such constitutively activated versions of 25 FSHR for the uses disclosed herein, as well as other uses.

#### C. Disease/Disorder Identification and/or Selection

[0077] As will be set forth in greater detail below, most preferably inverse agonists, partial agonists and agonists in the form of small molecule chemical compounds to the non-endogenous, constitutively activated GPCR can be identified by the methodologies of this 30 invention. Such compounds are ideal candidates as lead modulators in drug discovery programs for treating diseases or disorders associated with a particular receptor. The ability to directly

identify such compounds to the GPCR, in the absence of use of the receptor's endogenous ligand, allows for the development of pharmaceutical compositions.

[0078] Preferably, in situations where it is unclear what disease or disorder may be associated with a receptor; the DNA sequence of the GPCR is used to make a probe for (a) dot-blot analysis against tissue-mRNA, and/or (b) RT-PCR identification of the expression of the receptor in tissue samples. The presence of a receptor in a tissue source, or a diseased tissue, or the presence of the receptor at elevated concentrations in diseased tissue compared to a normal tissue, can be preferably utilized to identify a correlation with a treatment regimen, including but not limited to, a disease associated with that disease. Receptors can equally well be localized to regions of organs by this technique. Based on the known functions of the specific tissues to which the receptor is localized, the putative functional role of the receptor can be deduced.

#### D. Screening of Candidate compounds

##### 1. Generic GPCR screening assay techniques

[0079] When a G protein receptor becomes constitutively active, it binds to a G protein (e.g., Gq, Gs, Gi, Gz, Go) and stimulates the binding of GTP to the G protein. The G protein then acts as a GTPase and slowly hydrolyzes the GTP to GDP, whereby the receptor, under normal conditions, becomes deactivated. However, constitutively activated receptors continue to exchange GDP to GTP. A non-hydrolyzable analog of GTP, [<sup>35</sup>S]GTPγS, can be used to monitor enhanced binding to membranes which express constitutively activated receptors. It is reported that [<sup>35</sup>S]GTPγS can be used to monitor G protein coupling to membranes in the absence and presence of ligand. An example of this monitoring, among other examples well-known and available to those in the art, was reported by Traynor and Nahorski in 1995. The preferred use of this assay system is for initial screening of candidate compounds because the system is generically applicable to all G protein-coupled receptors regardless of the particular G protein that interacts with the intracellular domain of the receptor.

##### 2. Specific GPCR screening assay techniques

[0080] Once candidate compounds are identified using the "generic" G protein-coupled receptor assay (i.e., an assay to select compounds that are agonists, partial agonists, or inverse agonists), further screening to confirm that the compounds have interacted at the receptor site is preferred. For example, a compound identified by the "generic" assay may not bind to the receptor, but may instead merely "uncouple" the G protein from the intracellular domain.

*Gs, Gz and Gi.*

[0081] Gs stimulates the enzyme adenylyl cyclase. Gi (and Gz and Go), on the other hand, inhibit this enzyme. Adenylyl cyclase catalyzes the conversion of ATP to cAMP; thus, constitutively activated GPCRs that couple the Gs protein are associated with increased cellular levels of cAMP. On the other hand, constitutively activated GPCRs that couple Gi (or Gz, Go) protein are associated with decreased cellular levels of cAMP. *See, generally*, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3<sup>rd</sup> Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Thus, assays that detect cAMP can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to the receptor (*i.e.*, such a compound would decrease the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a most preferred approach relies upon the use of anti-cAMP antibodies in an ELISA-based format. Another type of assay that can be utilized is a second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or transcription factor (CREB) that then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, *e.g.*,  $\beta$ -galactosidase or luciferase. Thus, a constitutively activated Gs-linked receptor causes the accumulation of cAMP that then activates the gene and expression of the reporter protein. The reporter protein such as  $\beta$ -galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995).

*Go and Gq.*

[0082] Gq and Go are associated with activation of the enzyme phospholipase C, which in turn hydrolyzes the phospholipid PIP<sub>2</sub>, releasing two intracellular messengers: diacylglycerol (DAG) and initol 1,4,5-triphoisphate (IP<sub>3</sub>). Increased accumulation of IP<sub>3</sub> is associated with activation of Gq- and Go-associated receptors. *See, generally*, "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3<sup>rd</sup> Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Assays that detect IP<sub>3</sub> accumulation can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to a Gq- or Go-associated receptor (*i.e.*, such a compound would decrease the levels of IP<sub>3</sub>). Gq-associated receptors can also be examined using an AP1 reporter assay in that Gq-dependent phospholipase C causes activation of genes containing AP1 elements; thus, activated Gq-associated receptors will evidence an increase in the

expression of such genes, whereby inverse agonists thereto will evidence a decrease in such expression, and agonists will evidence an increase in such expression. Commercially available assays for such detection are available.

### 3. Ligand-Based Confirmation Assays

5 [0083] The candidate compounds directly identified using the techniques (or equivalent techniques) above are then, most preferably, verified using a ligand-based verification assay, such as the one set forth in the protocol of Example 8. The importance here is that the candidate compound be directly identified; subsequent confirmation, if any, using the endogenous ligand, is merely to confirm that the directly identified candidate compound has targeted the receptor.

### 10 4. GPCR Fusion Protein

[0084] The use of a non-endogenous, constitutively activated GPCR, for use in screening of candidate compounds for the direct identification of inverse agonists, agonists and partial agonists, provides an interesting screening challenge in that, by definition, the receptor is active even in the absence of an endogenous ligand bound thereto. Thus, in order to differentiate  
15 between, *e.g.*, the non-endogenous receptor in the presence of a candidate compound and the non-endogenous receptor in the absence of that compound, with an aim of such a differentiation to allow for an understanding as to whether such compound may be an inverse agonist, agonist, partial agonist or has no affect on such a receptor, it is preferred that an approach be utilized that can enhance such differentiation. A preferred approach is the use of a GPCR Fusion Protein.

20 [0085] Generally, once it is determined that a non-endogenous GPCR has been constitutively activated using the assay techniques set forth above (as well as others), it is possible to determine the predominant G protein that couples with the endogenous GPCR. Coupling of the G protein to the GPCR provides a signaling pathway that can be assessed. Because it is most preferred that screening take place by use of a mammalian expression system,  
25 such a system will be expected to have endogenous G protein therein. Thus, by definition, in such a system, the non-endogenous, constitutively activated GPCR will continuously signal. In this regard, it is preferred that this signal be enhanced such that in the presence of, *e.g.*, an inverse agonist to the receptor, it is more likely that it will be able to more readily differentiate, particularly in the context of screening, between the receptor when it is contacted with the inverse  
30 agonist.

[0086] The GPCR Fusion Protein is intended to enhance the efficacy of G protein coupling with the non-endogenous GPCR. The GPCR Fusion Protein is preferred for screening with a non-endogenous, constitutively activated GPCR because such an approach increases the signal that is most preferably utilized in such screening techniques. This is important in facilitating a significant "signal to noise" ratio; such a significant ratio is preferred for the screening of candidate compounds as disclosed herein.

[0087] The construction of a construct useful for expression of a GPCR Fusion Protein is within the purview of those having ordinary skill in the art. Commercially available expression vectors and systems offer a variety of approaches that can fit the particular needs of an investigator. The criteria of importance for such a GPCR Fusion Protein construct is that the endogenous GPCR sequence and the G protein sequence both be in-frame (preferably, the sequence for the endogenous GPCR is upstream of the G protein sequence) and that the "stop" codon of the GPCR must be deleted or replaced such that upon expression of the GPCR, the G protein can also be expressed. The GPCR can be linked directly to the G protein, or there can be spacer residues between the two (preferably, no more than about 12, although this number can be readily ascertained by one of ordinary skill in the art). Use of a spacer is preferred (based upon convenience) in that some restriction sites that are not used will, effectively, upon expression, become a spacer. Most preferably, the G protein that couples to the non-endogenous GPCR will have been identified prior to the creation of the GPCR Fusion Protein construct. Because there are only a few G proteins that have been identified, it is preferred that a construct comprising the sequence of the G protein (*i.e.*, a universal G protein construct) be available for insertion of an endogenous GPCR sequence therein; this provides for efficiency in the context of large-scale screening of a variety of different endogenous GPCRs having different sequences.

**E. Co-transfection of a Target Gi Coupled GPCR with a Signal-Enhancer Gs Coupled GPCR (cAMP Based Assays)**

[0088] A Gi coupled receptor is known to inhibit adenylyl cyclase, and, therefore, decrease the level of cAMP production, which can make assessment of cAMP levels challenging. An effective technique in measuring the decrease in production of cAMP as an indication of constitutive activation of a receptor that predominantly couples Gi upon activation can be accomplished by co-transfecting a signal enhancer, *e.g.*, a non-endogenous, constitutively activated receptor that predominantly couples with Gs upon activation (*e.g.*, TSHR-A623I, disclosed below), with the Gi linked GPCR. As is apparent, constitutive activation of a Gs

coupled receptor can be determined based upon an increase in production of cAMP. Constitutive activation of a Gi coupled receptor leads to a decrease in production cAMP. Thus, the co-transfection approach is intended to advantageously exploit these "opposite" affects. For example, co-transfection of a non-endogenous, constitutively activated Gs coupled receptor (the "signal enhancer") with the endogenous Gi coupled receptor (the "target receptor") provides a baseline cAMP signal (*i.e.*, although the Gi coupled receptor will decrease cAMP levels, this "decrease" will be relative to the substantial increase in cAMP levels established by constitutively activated Gs coupled signal enhancer). By then co-transfecting the signal enhancer with a constitutively activated version of the target receptor, cAMP would be expected to further decrease (relative to base line) due to the increased functional activity of the Gi target (*i.e.*, which decreases cAMP).

[0089] Screening of candidate compounds using a cAMP based assay can then be accomplished, with two provisos: first, relative to the Gi coupled target receptor, "opposite" effects will result, *i.e.*, an inverse agonist of the Gi coupled target receptor will increase the measured cAMP signal, while an agonist of the Gi coupled target receptor will decrease this signal; second, as would be apparent, candidate compounds that are directly identified using this approach should be assessed independently to ensure that these do not target the signal enhancing receptor (this can be done prior to or after screening against the co-transfected receptors).

#### F. Medicinal Chemistry

[0090] Generally, but not always, direct identification of candidate compounds is preferably conducted in conjunction with compounds generated via combinatorial chemistry techniques, whereby thousands of compounds are randomly prepared for such analysis. Generally, the results of such screening will be compounds having unique core structures; thereafter, these compounds are preferably subjected to additional chemical modification around a preferred core structure(s) to further enhance the medicinal properties thereof. Such techniques are known to those in the art and will not be addressed in detail in this patent document.

#### G. Pharmaceutical compositions

[0091] Candidate compounds selected for further development can be formulated into pharmaceutical compositions using techniques well known to those in the art. Suitable pharmaceutically-acceptable carriers are available to those in the art; for example, see

Remington's Pharmaceutical Sciences, 16<sup>th</sup> Edition, 1980, Mack Publishing Co., (Oslo et al., eds.).

#### H. Other Utility

[0092] Although a preferred use of the non-endogenous version of the known FSHR disclosed herein may be for the direct identification of candidate compounds as inverse agonists, agonists partial agonists or antagonist (preferably for use as pharmaceutical agents), these versions of known FSHR can also be utilized in research settings. For example, *in vitro* and *in vivo* systems incorporating GPCRs can be utilized to further elucidate and better understand the roles these receptors play in the human condition, both normal and diseased, as well as understanding the role of constitutive activation as it applies to understanding the signaling cascade. Other uses of the disclosed receptors will become apparent to those in the art based upon, *inter alia*, a review of this patent document.

#### EXAMPLES

[0093] The following examples are presented for purposes of elucidation, and not limitation, of the present invention. While specific nucleic acid and amino acid sequences are disclosed herein, those of ordinary skill in the art are credited with the ability to make minor modifications to these sequences while achieving the same or substantially similar results reported below. The traditional approach to application or understanding of sequence cassettes from one sequence to another (*e.g.* from rat receptor to human receptor or from human receptor A to human receptor B) is generally predicated upon sequence alignment techniques whereby the sequences are aligned in an effort to determine areas of commonality. The mutational approaches disclosed herein do not rely upon a sequence alignment approach but are instead based upon an algorithmic approach and a positional distance from a conserved proline residue located within the TM6 region of GPCRs. Once this approach is secured, those in the art are credited with the ability to make minor modifications thereto to achieve substantially the same results (*i.e.*, constitutive activation) disclosed herein. Such modified approaches are considered within the purview of this disclosure.

**Example 1****PREPARATION OF ENDOGENOUS GPCR: FSHR**

[0094] The 5' half portion of FSHR was cloned by PCR using testis cDNA as template and the following oligonucleotides:

5 5'-ATCACCATGGCCCTGCTCCTGGTCTCTTTG-3' (SEQ.ID.NO.:3)

5'-TGCCTTAAAATAGATTTGTTGCAAATTGGA-3' (SEQ.ID.NO.:4).

[0095] The 3' half of FSHR was cloned by PCR using genomic DNA as template and the following oligonucleotides:

5'-CTCTGAGCTTCATCCAATTGCAACAAATC-3' (SEQ.ID.NO.:5)

10 5'-TGTGAATTCGTTTTGGGCTAAATGACTTAGAGGGAC-3' (SEQ.ID.NO.:6).

[0096] The 900bp fragment of 5' PCR and the 1.24 Kb 3' PCR fragment were then used as co-template to perform secondary PCR using kinased oligonucleotides with SEQ.ID.NO.:3 and SEQ.ID.NO.:6.

[0097] PCR was performed using rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25μM of each oligonucleotide, and 0.2mM of each four (4) nucleotides. The cycle condition was 30 cycles of 94°C for 1 min., 65°C for 1 min., and 72°C for 2 min. and 30 sec. The 2.1 kb PCR fragment was then cloned into EcoRV-EcoRI site of CMVp expression vector. See, SEQ.ID.NO.:1 for the nucleic acid sequence and SEQ.ID.NO.:2 for the putative amino acid sequence.

20 **Example 2**

**PREPARATION OF NON-ENDOGENOUS VERSIONS OF GPCR: FSHR**

[0098] Those skilled in the art are credited with the ability to select techniques for mutation of a nucleic acid sequence. Presented below are approaches utilized to create non-endogenous versions of human FSHR disclosed above. The mutations disclosed below are based upon an algorithmic approach whereby the 16<sup>th</sup> amino acid (located in the IC3 region of the GPCR) from a conserved proline (or an endogenous, conservative substitution therefore) residue (located in the TM6 region of the GPCR, near the TM6/IC3 interface) is mutated, preferably to an alanine, histamine, arginine or lysine amino acid residue, most preferably to a lysine amino acid residue.

[0099] Preparation of non-endogenous human versions of FSHR were accomplished by using QuikChange™ Site-Directed™ Mutagenesis Kit (Stratagene, according to manufacturer's instructions). Endogenous FSHR (*see* Example 1 above) was preferably used as a template and the following oligonucleotides were used to create non-endogenous versions of FSHR. For convenience, the codon mutations incorporated into the human FSHR are in standard form in Table B below.

TABLE B

Codon Mutation	5' Primer (SEQ.ID.NO.)	3' Primer (SEQ.ID.NO.)
A376V	TTATCAGCATCCTGG TCATCACTGGGAACA T (7)	ATGTTCCCAGTGA TGACCAGGATGCT GATAA (8)
V457A	CCAGTGAGCTGTCAG CCTACACTCTGACAG C (9)	GCTGTCAGAGTGT AGGCTGACAGCTC ACTGG (10)
L460R	TGTCAGTCTACACTCG GACAGCTATCACCTT (11)	AAGGTGATAGCTG TCCGAGTGTAGAC TGACA (12)
D567G	TGTCCTCCTCTAGTG GCACCAGGATCGCCA A (13)	TTGGCGATCCTGG TGCCACTAGAGGA GGACA (14)
A571K	AGTGACACCAGGATC AAGAAGCGCATGGCC ATG (15)	CATGGCCATGCGC TTCTTGATCCTGG TGTCAT (16)
D581G	TGCTCATCTTCACTG GCTTCCTCTGCATGG C (17)	GCCATGCAGAGGA AGCCAGTGAAGAT GAGCA (18)
C620Y	ACCCCATCAACTCCT ATGCCAACCCCTTCC T (19)	AGGAAGGGGTTG GCATAGGAGTTGA TGGGGT (20)

[0101] The non-endogenous versions of human FSHR were then sequenced and the derived and verified nucleic acid and amino acid sequences are listed in the accompanying "Sequence Listing" appendix to this patent document, as summarized in Table C below:

Codon Mutation	Nucleic Acid Sequence Listing	Amino Acid Sequence Listing
A376V	SEQ.ID.NO.:21	SEQ.ID.NO.:22
V457A	SEQ.ID.NO.:23	SEQ.ID.NO.:24
L460R	SEQ.ID.NO.:25	SEQ.ID.NO.:26
D567G	SEQ.ID.NO.:27	SEQ.ID.NO.:28
A571K	SEQ.ID.NO.:29	SEQ.ID.NO.:30
D581G	SEQ.ID.NO.:31	SEQ.ID.NO.:32
C620Y	SEQ.ID.NO.:33	SEQ.ID.NO.:34

**[0102]** Assessment of constitutive activity of the non-endogenous versions of human FSHR was then accomplished. *See*, Example 4 below:

### 5 **Example 3** **RECEPTOR EXPRESSION**

**[0103]** Although a variety of cells are available to the art for the expression of proteins, it is most preferred that mammalian cells be utilized. The primary reason for this is predicated upon practicalities, *i.e.*, utilization of, *e.g.*, yeast cells for the expression of a GPCR, while possible, introduces into the protocol a non-mammalian cell which may not (indeed, in the case of yeast, does not) include the receptor-coupling, genetic-mechanism and secretory pathways that have evolved for mammalian systems – thus, results obtained in non-mammalian cells, while of potential use, are not as preferred as that obtained from mammalian cells. Of the mammalian cells, COS-7, 293 and 293T cells are particularly preferred, although the specific mammalian cell utilized can be predicated upon the particular needs of the artisan.

#### **a. Transient Transfection**

**[0104]** On day one,  $4 \times 10^6$  of 293 cells were plated out. On day two, two reaction tubes were prepared (the proportions to follow for each tube are per plate): tube A was prepared by mixing 4 $\mu$ g DNA (*e.g.*, pCMV vector; pCMV vector with receptor cDNA, etc.) in 0.5 ml serum free DMEM (Gibco BRL); tube B was prepared by mixing 24 $\mu$ l lipofectamine (Gibco BRL) in 0.5ml serum free DMEM. Tubes A and B were admixed by inversions (several times), followed by incubation at room temperature for 30–45min. The admixture is referred to as the “transfection mixture”. Plated 293 cells were washed with 1XPBS, followed by addition of 5 ml serum free DMEM. 1 ml of the transfection mixture were added to the cells, followed by incubation for 4hrs at 37°C/5% CO<sub>2</sub>. The transfection mixture was removed by aspiration,

followed by the addition of 10ml of DMEM/10% Fetal Bovine Serum. Cells were incubated at 37°C/5% CO<sub>2</sub>. After 48hr incubation, cells were harvested and utilized for analysis.

b. **Stable Cell Lines**

[0105] Approximately  $12 \times 10^6$  293 cells are plated on a 15cm tissue culture plate. Grown in DME High Glucose Medium containing ten percent fetal bovine serum and one percent sodium pyruvate, L-glutamine, and anti-biotics. Twenty-four hours following plating of 293 cells (or to ~80% confluency), the cells are transfected using 12µg of DNA. The 12µg of DNA is combined with 60µl of lipofectamine and 2mL of DME High Glucose Medium without serum. The medium is aspirated from the plates and the cells are washed once with medium without serum. The DNA, lipofectamine, and medium mixture are added to the plate along with 10mL of medium without serum. Following incubation at 37 degrees Celsius for four to five hours, the medium is aspirated and 25ml of medium containing serum is added. Twenty-four hours following transfection, the medium is aspirated again, and fresh medium with serum is added. Forty-eight hours following transfection, the medium is aspirated and medium with serum is added containing geneticin (G418 drug) at a final concentration of 500µg/mL. The transfected cells will undergo selection for positively transfected cells containing the G418 resistant gene. The medium is replaced every four to five days as selection occurs. During selection, cells are grown to create stable pools, or split for stable clonal selection.

20 **Example 4**  
**ASSAYS FOR DETERMINATION OF CONSTITUTIVE ACTIVITY**  
**OF NON-ENDOGENOUS FSHR**

[0106] A variety of approaches are available for assessment of constitutive activity of the non-endogenous versions of human FSHR. The following are illustrative; those of ordinary skill in the art are credited with the ability to determine those techniques that are preferentially beneficial for the needs of the artisan.

30 1. **Membrane Binding Assays: [<sup>35</sup>S]GTPγS Assay**

[0107] When a G protein-coupled receptor is in its active state, either as a result of ligand binding or constitutive activation, the receptor couples to a G protein and stimulates the release of GDP and subsequent binding of GTP to the G protein. The alpha subunit of the G protein-receptor complex acts as a GTPase and slowly hydrolyzes the GTP to GDP, at which

point the receptor normally is deactivated. Constitutively activated receptors continue to exchange GDP for GTP. The non-hydrolyzable GTP analog, [<sup>35</sup>S]GTPγS, can be utilized to demonstrate enhanced binding of [<sup>35</sup>S]GTPγS to membranes expressing constitutively activated receptors. The advantage of using [<sup>35</sup>S]GTPγS binding to measure constitutive activation is that:

- 5 (a) it is generically applicable to all G protein-coupled receptors; (b) it is proximal at the membrane surface making it less likely to pick-up molecules which affect the intracellular cascade.

[0108] The assay utilizes the ability of G protein coupled receptors to stimulate [<sup>35</sup>S]GTPγS binding to membranes expressing the relevant receptors. The assay can, therefore, be  
10 used in the direct identification method to screen candidate compounds to known, orphan and constitutively activated G protein-coupled receptors. The assay is generic and has application to drug discovery at all G protein-coupled receptors.

[0109] The [<sup>35</sup>S]GTPγS is incubated in 20 mM HEPES and between 1 and about 20mM MgCl<sub>2</sub> (this amount can be adjusted for optimization of results, although 20mM is preferred) pH  
15 7.4, binding buffer with between about 0.3 and about 1.2 nM [<sup>35</sup>S]GTPγS (this amount can be adjusted for optimization of results, although 1.2 is preferred ) and 12.5 to 75 μg membrane protein (e.g. 293 cells expressing the Gs Fusion Protein; this amount can be adjusted for optimization) and 10 μM GDP (this amount can be changed for optimization) for 1 hour. Wheatgerm agglutinin beads (25 μl; Amersham) are then added and the mixture incubated for  
20 another 30 minutes at room temperature. The tubes are then centrifuged at 1500 x g for 5 minutes at room temperature and then counted in a scintillation counter.

## 2. Cell-based cAMP Detection Assay

[0110] In the following assay, a 96-well Adenylyl Cyclase Activation Flashplate was used (NEN: #SMP004A). First, 50ul of the standards for the assay were added to the plate, in  
25 duplicate, ranging from concentrations of 50pmol to zero pmol cAMP per well. The standard cAMP (NEN: #SMP004A) was reconstituted in water, and serial dilutions were made using 1xPBS (Irvine Scientific: #9240). Next, 50ul of the stimulation buffer (NEN: #SMP004A) was added to all wells. Various final concentrations used range from 1uM up to 1mM. Adenosine 5'-triphosphate, ATP, (Research Biochemicals International: #A-141) and Adenosine 5'-diphosphate, ADP, (Sigma: #A2754) were used in the assay. Next, the 293 cells transfected with  
30 12ug (per 150mm tissue culture plate) of the respective cDNA (CMV or FSHR) were harvested

24 hours post-transfection. The media was aspirated and the cells washed once with 1xPBS. Then 5ml of 1xPBS was added to the cells along with 3ml of cell dissociation buffer (Sigma: #C-1544). The detached cells were transferred to a centrifuge tube and centrifuged at room temperature for five minutes. The supernatant was removed and the cell pellet was resuspended in an appropriate amount of 1xPBS to obtain a final concentration of  $2 \times 10^6$  cells per milliliter.

[0111] The plate was incubated on a shaker for 15 minutes at room temperature. The detection buffer containing the tracer cAMP was prepared. In 11ml of detection buffer (NEN: #SMP004A), 50ul (equal to 1uCi) of [ $^{125}$ I]cAMP (NEN: #SMP004A) was added. Following incubation, 50ul of this detection buffer containing tracer cAMP was added to each well. The plate was placed on a shaker and incubated at room temperature for two hours. Finally, the solution from the wells of the plate were aspirated and the flashplate was counted using the Wallac MicroBeta plate reader.

[0112] Reference is made to Figure 1. Figure 1 depicts the results of a second messenger cell-based cyclic AMP assay providing comparative results for constitutive signaling of endogenous FSHR ("FSHRwt"), non-endogenous versions of FSHR ("L460R", "A376V", "V457A", "L460R", "D567G", "A571K", "D581G", and "C620Y") and a control vector ("CMV"). This data evidences that the L460R version of FSHR is constitutively activated by about a ten (10) fold increase in cAMP production.

### 3. Alpha Screen

[0113] The media from Example 3(b) above was aspirated and rinsed 1X with PBS (5-10ml/flask). 10-20mls of PBS was then added to each flask and let sit for 2-5 minute. The cells were then pipetted off into conocal tubes for spinning for 5 minutes at 1500rpm. PBS was apriated and re-suspended with Stimulation Buffer (1X HBSS, 0.5mM IBMX, 5mM Hepes and 011% BSA). 2 % DMSO diluted the Hepes Buffer and 10ul/well of cells at 15,000cells/well were then added to the wells and incubated for 30 minutes. 5ul/well of cAMP Acceptor Beads (Perkin Elmer Product No. 6760600R) for a final concentration of 15ug/ml. The wells were then covered and left to incubate for two hours at room temperature. 5ul of Assay Reaction Mixture was added. The Assay Reaction Mixture was prepared by mixing the Donor Bead (Perkin Elmer Product No. 6760600R) with a final concentration of 20ug/ml, Biotinylated cAMP Mix (Perkin Elmer Product No. 6760600R) with a final concentration of 10nM, and Lysis Buffer (5mM Hepes

and 0.18% Igapel). The wells were then covered and incubated for two hours at room temperature. Following incubation, the wells were read on Alpha Quest and measured for light units. The light unit was then converted to pmol cAMP/well by taking the cAMP concentration and determining the pmol/well of cAMP and using the linear regression function found on GraphPad Prism version 3.00 for Windows, GraphPad Software, San Diego California USA, the light units were converted to pmol cAMP/well.

[0114] Reference is made to Figure 2. Figure 2 depicts the results of a second messenger cAMP accumulation assay providing comparative results for constitutive signaling of endogenous FSHR ("WT"), non-endogenous version of FSHR ("L460R") and a control vector ("CMV"). This data further evidences that the L460R version of FSHR is constitutively activated by about a twenty-eight (28) fold increase in cAMP production.

#### 4. Cell-Based cAMP for Gi Coupled Target GPCRs

[0115] TSHR is a Gs coupled GPCR that causes the accumulation of cAMP upon activation. TSHR is constitutively activated by mutating amino acid residue 623 (*i.e.*, changing an alanine residue to an isoleucine residue). A G<sub>i</sub> coupled receptor is expected to inhibit adenylyl cyclase, and, therefore, decrease the level of cAMP production, which can make assessment of cAMP levels challenging. An effective technique for measuring the decrease in production of cAMP as an indication of constitutive activation of a G<sub>i</sub> coupled receptor can be accomplished by co-transfecting, most preferably, non-endogenous, constitutively activated TSHR (TSHR-A623I) (or an endogenous, constitutively active G<sub>s</sub> coupled receptor) as a "signal enhancer" with a G<sub>i</sub> linked target GPCR to establish a baseline level of cAMP. Upon creating a non-endogenous version of the G<sub>i</sub> coupled receptor, this non-endogenous version of the target GPCR is then co-transfected with the signal enhancer, and it is this material that can be used for screening. We will utilize such approach to effectively generate a signal when a cAMP assay is used; this approach is preferably used in the direct identification of candidate compounds against G<sub>i</sub> coupled receptors. It is noted that for a G<sub>i</sub> coupled GPCR, when this approach is used, an inverse agonist of the target GPCR will increase the cAMP signal and an agonist will decrease the cAMP signal.

[0116] On day one,  $2 \times 10^4$  293 cells is plated out. On day two, two reaction tubes are prepared (the proportions to follow for each tube are per plate): tube A is prepared by mixing 2 $\mu$ g DNA of each receptor transfected into the mammalian cells, for a total of 4 $\mu$ g DNA (*e.g.*, pCMV vector; pCMV vector with mutated TSHR (TSHR-A623I); TSHR-A623I and GPCR, etc.) in

1.2ml serum free DMEM (Irvine Scientific, Irvine, CA); tube B is prepared by mixing 120µl lipofectamine (Gibco BRL) in 1.2ml serum free DMEM. Tubes A and B are be admixed by inversions (several times), followed by incubation at room temperature for 30-45min. The admixture is referred to as the "transfection mixture". Plated 293 cells are washed with 1XPBS, followed by addition of 10ml serum free DMEM. 2.4ml of the transfection mixture is then added to the cells, followed by incubation for 4hrs at 37°C/5% CO<sub>2</sub>. The transfection mixture is removed by aspiration, followed by the addition of 25ml of DMEM/10% Fetal Bovine Serum. Cells are incubated at 37°C/5% CO<sub>2</sub>. After 24hr incubation, cells are then harvested and utilized for analysis.

10 [0117] A Flash Plate™ Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) is designed for cell-based assays, however, can be modified for use with crude plasma membranes depending on the need of the skilled artisan. The Flash Plate wells will contain a scintillant coating which also contains a specific antibody recognizing cAMP. The cAMP generated in the wells can be quantitated by a direct competition for binding of radioactive cAMP tracer to the cAMP antibody. The following serves as a brief protocol for the measurement of changes in cAMP levels in whole cells that express the receptors.

20 [0118] Transfected cells are harvested approximately twenty four hours after transient transfection. Media is carefully aspirated off and discarded. 10ml of PBS is gently added to each dish of cells followed by careful aspiration. 1ml of Sigma cell dissociation buffer and 3ml of PBS are added to each plate. Cells are pipetted off the plate and the cell suspension is collected into a 50ml conical centrifuge tube. Cells are then centrifuged at room temperature at 1,100 rpm for 5 min. The cell pellet is carefully re-suspended into an appropriate volume of PBS (about 3ml/plate). The cells are counted using a hemocytometer and additional PBS is added to give the appropriate number of cells (with a final volume of about 50µl/well).

25 [00100] cAMP standards and Detection Buffer (comprising 1 µCi of tracer [<sup>125</sup>I cAMP (50 µl] to 11 ml Detection Buffer) are prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer should be prepared fresh for screening and contained 50µl of Stimulation Buffer, 3µl of test compound (12µM final assay concentration) and 50µl cells, Assay Buffer can be stored on ice until utilized. The assay can be initiated by addition of 30 50µl of cAMP standards to appropriate wells followed by addition of 50µl of PBSA to wells H-11 and H12. Fifty µl of Stimulation Buffer is added to all wells. Selected compounds (e.g., FSH) is added to appropriate wells using a pin tool capable of dispensing 3µl of compound solution,

with a final assay concentration of 12µM test compound and 100µl total assay volume. The cells are added to the wells and incubated for 60 min at room temperature. 100µl of Detection Mix containing tracer cAMP is then added to the wells. Plates are incubated additional 2 hours followed by counting in a Wallac MicroBeta scintillation counter. Values of cAMP/well are then extrapolated from a standard cAMP curve which is contained within each assay plate.

## 5. Reporter-Based Assays

### a. CRE-LUC Reporter Assay (Gs-associated receptors)

[0119] 293 and 293T cells are plated-out on 96 well plates at a density of  $2 \times 10^4$  cells per well and transfected using Lipofectamine Reagent (BRL) the following day according to manufacturer instructions. A DNA/lipid mixture is prepared for each 6-well transfection as follows: 260ng of plasmid DNA in 100µl of DMEM is gently mixed with 2µl of lipid in 100µl of DMEM (the 260ng of plasmid DNA consisted of 200ng of a 8xCRE-Luc reporter plasmid, 50ng of pCMV comprising endogenous receptor or non-endogenous receptor or pCMV alone, and 10ng of a GPRS expression plasmid (GPRS in pcDNA3 (Invitrogen)). The 8xCRE-Luc reporter plasmid is prepared as follows: vector SRIF-β-gal was obtained by cloning the rat somatostatin promoter (-71/+51) at BglV-HindIII site in the pβgal-Basic Vector (Clontech). Eight (8) copies of cAMP response element will be obtained by PCR from an adenovirus template AdpCF126CCRE8 (see, 7 *Human Gene Therapy* 1883 (1996)) and cloned into the SRIF-β-gal vector at the Kpn-BglV site, resulting in the 8xCRE-β-gal reporter vector. The 8xCRE-Luc reporter plasmid is generated by replacing the beta-galactosidase gene in the 8xCRE-β-gal reporter vector with the luciferase gene obtained from the pGL3-basic vector (Promega) at the HindIII-BamHI site. Following 30 min. incubation at room temperature, the DNA/lipid mixture is diluted with 400 µl of DMEM and 100µl of the diluted mixture is added to each well. 100 µl of DMEM with 10% FCS is added to each well after a 4hr incubation in a cell culture incubator. The following day the transfected cells are changed with 200 µl/well of DMEM with 10% FCS. Eight (8) hours later, the wells are changed to 100 µl /well of DMEM without phenol red, after one wash with PBS. Luciferase activity is measured the next day using the LucLite™ reporter gene assay kit (Packard) following manufacturer instructions and read on a 1450 MicroBeta™ scintillation and luminescence counter (Wallac).

### 30 b. AP1 reporter assay (Gq-associated receptors)

[0120] A method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing AP1 elements in their

promoter. A Pathdetect™ AP-1 cis-Reporting System (Stratagene, Catalogue # 219073) can be utilized following the protocol set forth above with respect to the CREB reporter assay, except that the components of the calcium phosphate precipitate were 410 ng pAP1-Luc, 80 ng pCMV-receptor expression plasmid, and 20 ng CMV-SEAP.

5           c.       **SRF-LUC Reporter Assay (Gq- associated receptors)**

[0121]       One method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing serum response factors in their promoter. A Pathdetect™ SRF-Luc-Reporting System (Stratagene) can be utilized to assay for Gq coupled activity in, *e.g.*, COS7 cells. Cells are transfected with the plasmid components of the system and the indicated expression plasmid encoding endogenous or non-endogenous GPCR using a Mammalian Transfection™ Kit (Stratagene, Catalogue #200285) according to the manufacturer's instructions. Briefly, 410 ng SRF-Luc, 80 ng pCMV-receptor expression plasmid and 20 ng CMV-SEAP (secreted alkaline phosphatase expression plasmid; alkaline phosphatase activity is measured in the media of transfected cells to control for variations in transfection efficiency between samples) are combined in a calcium phosphate precipitate as per the manufacturer's instructions. Half of the precipitate is equally distributed over 3 wells in a 96-well plate, kept on the cells in a serum free media for 24 hours. The last 5 hours the cells are incubated with 1µM Angiotensin, where indicated. Cells are then lysed and assayed for luciferase activity using a Lucite™ Kit (Packard, Cat. # 6016911) and "Trilux 1450 Microbeta" liquid scintillation and luminescence counter (Wallac) as per the manufacturer's instructions. The data can be analyzed using GraphPad Prism™ 2.0a (GraphPad Software Inc.).

          d.       **Intracellular IP<sub>3</sub> Accumulation Assay (G<sub>q</sub>-associated receptors)**

[00101]       On day 1, cells comprising the receptors (endogenous and/or non-endogenous) can be plated onto 24 well plates, usually 1x10<sup>5</sup> cells/well (although this number can be optimized. On day 2 cells can be transfected by firstly mixing 0.25µg DNA in 50 µl serum free DMEM/well and 2 µl lipofectamine in 50 µl serumfree DMEM/well. The solutions are gently mixed and incubated for 15-30 min at room temperature. Cells are washed with 0.5 ml PBS and 400 µl of serum free media is mixed with the transfection media and added to the cells. The cells are then incubated for 3-4 hrs at 37°C/5%CO<sub>2</sub> and then the transfection media is removed and replaced with 1ml/well of regular growth media. On day 3 the cells are labeled with <sup>3</sup>H-myo-inositol. Briefly, the media is removed and the cells are washed with 0.5 ml PBS. Then 0.5 ml inositol-free/serum free media (GIBCO BRL) is added/well with 0.25 µCi of <sup>3</sup>H-myo-inositol/ well and the cells are incubated for 16-18 hrs o/n at 37°C/5%CO<sub>2</sub>. On Day 4 the cells are washed with 0.5

ml PBS and 0.45 ml of assay medium is added containing inositol-free/serum free media 10  $\mu$ M pargyline 10 mM lithium chloride or 0.4 ml of assay medium and 50  $\mu$ l of 10x ketanserin (ket) to final concentration of 10  $\mu$ M. The cells are then incubated for 30 min at 37°C. The cells are then washed with 0.5 ml PBS and 200  $\mu$ l of fresh/ice cold stop solution (1M KOH; 18 mM Na-borate; 3.8 mM EDTA) is added/well. The solution is kept on ice for 5-10 min or until cells were lysed and then neutralized by 200  $\mu$ l of fresh/ice cold neutralization sol. (7.5 % HCL). The lysate is then transferred into 1.5 ml eppendorf tubes and 1 ml of chloroform/methanol (1:2) is added/tube. The solution is vortexed for 15 sec and the upper phase is applied to a Biorad AG1-X8™ anion exchange resin (100-200 mesh). Firstly, the resin is washed with water at 1:1.25 W/V and 0.9 ml of upper phase is loaded onto the column. The column is washed with 10 mls of 5 mM myo-inositol and 10 ml of 5 mM Na-borate/60mM Na-formate. The inositol tris phosphates are eluted into scintillation vials containing 10 ml of scintillation cocktail with 2 ml of 0.1 M formic acid/ 1 M ammonium formate. The columns are regenerated by washing with 10 ml of 0.1 M formic acid/3M ammonium formate and rinsed twice with dd H<sub>2</sub>O and stored at 4°C in water.

15

**Example 5****FUSION PROTEIN PREPARATION****a. GPCR:Gs Fusion Construct**

20 [0122] The design of the constitutively activated GPCR-G protein fusion construct can be accomplished as follows: both the 5' and 3' ends of the rat G protein Gs $\alpha$  (long form; Itoh, H. et al., 83 *PNAS* 3776 (1986)) are engineered to include a HindIII (5'-AAGCTT-3') sequence thereon. Following confirmation of the correct sequence (including the flanking HindIII sequences), the entire sequence is shuttled into pcDNA3.1(-) (Invitrogen, cat. no. V795-20) by subcloning using the HindIII restriction site of that vector. The correct orientation for the G $\alpha$  sequence is determined after subcloning into pcDNA3.1(-). The modified pcDNA3.1(-) containing the rat G $\alpha$  gene at HindIII sequence is then verified; this vector is now available as a "universal" G $\alpha$  protein vector. The pcDNA3.1(-) vector contains a variety of well-known restriction sites upstream of the HindIII site, thus beneficially providing the ability to insert, upstream of the Gs protein, the coding sequence of an endogenous, constitutively active GPCR. This same approach can be utilized to create other "universal" G protein vectors, and, of course, other commercially available or proprietary vectors known to the artisan can be utilized – the

30

important criteria is that the sequence for the GPCR be upstream and in-frame with that of the G protein.

[0123] PCR is then utilized to secure the respective receptor sequences for fusion within the Gs $\alpha$  universal vector disclosed above, using the following protocol for each: 100ng cDNA is added to separate tubes containing 2 $\mu$ L of each primer (sense and anti-sense), 3 $\mu$ L of 10mM dNTPs, 10 $\mu$ L of 10XTaqPlus™ Precision buffer, 1 $\mu$ L of TaqPlus™ Precision polymerase (Stratagene: #600211), and 80 $\mu$ L of water. Reaction temperatures and cycle times are as follows with cycle steps 2 through 4 were repeated 35 times: 94°C for 1 min; 94°C for 30 seconds; 62°C for 20 sec; 72°C 1 min 40sec; and 72° C 5 min . PCR product is then ran on a 1% agarose gel and purified. The purified product is then digested with XbaI and EcoRV and the desired inserts purified and ligated into the Gs universal vector at the respective restriction site. The positive clones are isolated following transformation and determined by restriction enzyme digest; expression using 293 cells is accomplished following the protocol set forth *infra*. Each positive clone for GPCR-Gs Fusion Protein is then sequenced to verify correctness.

#### 15 Gq(6 amino acid deletion)/Gi Fusion Construct

[0124] The design of a G<sub>q</sub> (del)/G<sub>i</sub> fusion construct can be accomplished as follows: the N-terminal six (6) amino acids (amino acids 2 through 7, having the sequence of TLESIM G $\alpha_q$ -subunit is deleted and the C-terminal five (5) amino acids, having the sequence EYNLV is replaced with the corresponding amino acids of the G $\alpha_i$  Protein, having the sequence DCGLF.

20 This fusion construct is obtained by PCR using the following primers:

5'-gatcaagcttcCATGGCGTGCTGCCTGAGCGAGGAG-3' (SEQ.ID.NO.:35) and  
5'-gatcggatccTTAGAACAGGCCGCGAGTCCTTCAGGTTTCAGCTGCAGGATGGTG-3'  
(SEQ.ID.NO.:36)

and Plasmid 63313 which contains the mouse G $\alpha_q$ -wild type version with a hemagglutinin tag as template. Nucleotides in lower caps are included as spacers.

[0125] TaqPlus Precision DNA polymerase (Stratagene) is utilized for the amplification by the following cycles, with steps 2 through 4 repeated 35 times: 95°C for 2 min; 95°C for 20 sec; 56°C for 20 sec; 72°C for 2 min; and 72°C for 7 min. The PCR product is cloned into a pCRII-TOPO vector (Invitrogen) and sequenced using the ABI Big Dye Terminator kit (P.E. Biosystems). Inserts from a TOPO clone containing the sequence of the fusion construct is shuttled into the expression vector pcDNA3.1(+) at the HindIII/BamHI site by a 2 step cloning process.

**EXAMPLE 6****TISSUE DISTRIBUTION OF THE DISCLOSED HUMAN GPCRS****5           A.     RT-PCR**

[0126]       RT-PCR is applied to confirm the expression and to determine the tissue distribution of human FSHR. Oligonucleotides utilized are FSHR-specific and the human multiple tissue cDNA panels (MTC, Clontech) as templates. Taq DNA polymerase (Stratagene) is utilized for the amplification in a 40µl reaction according to the manufacturer's instructions. 20µl of the reaction is loaded on a 1.5% agarose gel to analyze the RT-PCR products.

[0127]       Diseases and disorders related to receptors located in these tissues or regions include, but are not limited to, cardiac disorders and diseases (e.g. thrombosis, myocardial infarction; atherosclerosis; cardiomyopathies); kidney disease/disorders (e.g., renal failure; renal tubular acidosis; renal glycosuria; nephrogenic diabetes insipidus; cystinuria; polycystic kidney disease); eosinophilia; leukocytosis; leukopenia; ovarian cancer; sexual dysfunction; polycystic ovarian syndrome; pancreatitis and pancreatic cancer; irritable bowel syndrome; colon cancer; Crohn's disease; ulcerative colitis; diverticulitis; Chronic Obstructive Pulmonary Disease (COPD); Cystic Fibrosis; pneumonia; pulmonary hypertension; tuberculosis and lung cancer; Parkinson's disease; movement disorders and ataxias; learning and memory disorders; eating disorders (e.g., anorexia; bulimia, etc.); obesity; cancers; thymoma; myasthenia gravis; circulatory disorders; prostate cancer; prostatitis; kidney disease/disorders(e.g., renal failure; renal tubular acidosis; renal glycosuria; nephrogenic diabetes insipidus; cystinuria; polycystic kidney disease); sensorimotor processing and arousal disorders; obsessive-compulsive disorders; testicular cancer; priapism; prostatitis; hernia; endocrine disorders; sexual dysfunction; allergies; depression; psychotic disorders; migraine; reflux; schizophrenia; ulcers; bronchospasm; epilepsy; prostatic hypertrophy; anxiety; rhinitis; angina; and glaucoma. Accordingly, the methods of the present invention may also be useful in the diagnosis and/or treatment of these and other diseases and disorders.

30

**B.     Affymetrix GeneChip® Technology**

[0128]       Sequences from the public database are submitted to Affymetrix for the design and manufacture of microarrays containing oligonucleotides to monitor the expression levels of G protein-coupled receptors (GPCRs) using GeneChip® Technology. RNA samples are

amplified, labeled, hybridized to the microarray, and data analyzed according to manufacturer's instructions.

### Example 7

#### 5 Protocol: Direct Identification of Inverse Agonists and Agonists

##### A. Alpha Screen

[0129] The media from Example 3(b) above was aspirated and rinsed 1X with PBS (5-10ml/flask). 10-20mls of PBS was then added to each flask and let sit for 2-5 minute. The cells were then pipetted off into conocal tubes for spinning for 5 minutes at 1500rpm. PBS was  
10 apriated and re-suspended with Stimulation Buffer (1X HBSS, 0.5mM IBMX, 5mM Hepes and 0.11% BSA). 5µl/wll of Compound A diluted in Hepes Buffer and 10µl/well of cells at 15,000cells/well were then added to the wells and incubated for 30 minutes. 5µl/well of cAMP Acceptor Beads (Perkin Elmer Product No. 6760600R) for a final concentration of 15µg/ml. The wells were then covered and left to incubate for two hours at room temperature. 5µl of Assay  
15 Reaction Mixture was added. The Assay Reaction Mixture was prepared by mixing the Donor Bead (Perkin Elmer Product No. 6760600R) with a final concentration of 20µg/ml, Biotinylated cAMP Mix (Perkin Elmer Product No. 6760600R) with a final concentration of 10nM, and Lysis Buffer (5mM Hepes and 0.18% Igapel). The wells were then covered and incubated for two hours at room temperature. Following incubation, the wells were read on Alpha Quest and  
20 measured for light units. The light unit was then converted to pmol cAMP/well by taking the cAMP concentration and determining the pmol/well of cAMP and using the linear regretion function found on GraphPad Prism version 3.00 for Windows, GraphPad Software, San Diego California USA, the light units were converted to pmol cAMP/well.

[0130] Compound A is disclosed in U.S. Patent Numbers 6,235,755B1 and 6,423,723B1  
25 as falling within a class of compounds that have been shown to bind to the endogenous FSH receptor. Compound A used in this assay is chemically defined as 1-[(2-Oxo-6-pentyl-2H-pyran)-3-carbonyl]-piperidine-2-carboxylic acid-3-(9-ethylcarbazoly) amide. U.S. Patent Number 6,235,755B1 and 6,423,723B1 are incorporated herein by reference in its entirety.

[0131] Reference is made to Figure 3. Figure 3 depicts the results of cAMP  
30 accumulation of the endogenous FSHR ("WT") compared with the non-endogenous FSHR ("L460R") and a control vector ("CMV") in the presence of Compound A. Compound A binds to the WT receptor at an EC50 of about 3nM, while Compound A binds the L460R version of FSHR at about 7µM. This data evidences that Compound A has a better efficacy for the non-

endogenous, constitutively activated version of FHSR (L460) than the WT receptor. Therefore, the non-endogenous, constitutively activated version of FHSR can be used in a screening assay to screen for receptor compounds, including but not limited to, agonist, inverse agonist, partial agonist or antagonist.

5           **B.     <sup>35</sup>S]GTP $\gamma$ S Assay**

[0132]       Both endogenous and non-endogenous versions of human FSHR can be utilized for the direct identification of candidate compounds as, *e.g.*, inverse agonists. In some embodiments, a GPCR Fusion Protein, as disclosed above, can also be utilized with a non-endogenous, constitutively activated FSHR. When such a protein is used, intra-assay variation  
10 appears to be substantially stabilized, whereby an effective signal-to-noise ratio is obtained. This has the beneficial result of allowing for a more robust identification of candidate compounds. Thus, in some embodiments it is preferred that for direct identification, a FSHR Fusion Protein be used and that when utilized, the following assay protocols be utilized.

15       **Membrane Preparation**

[0133]       In some embodiments membranes comprising the constitutively active GPCR/Fusion Protein of interest and for use in the direct identification of candidate compounds as inverse agonists or agonists are preferably prepared as follows:

**a.     Materials**

20   [0134]       “Membrane Scrape Buffer” is comprised of 20mM HEPES and 10mM EDTA, pH 7.4; “Membrane Wash Buffer” is comprised of 20 mM HEPES and 0.1 mM EDTA, pH 7.4; “Binding Buffer” is comprised of 20mM HEPES, 100 mM NaCl, and 10 mM MgCl<sub>2</sub>, pH 7.4.

**b.     Procedure**

[0135]       All materials are kept on ice throughout the procedure. Firstly, the media is  
25 aspirated from a confluent monolayer of cells, followed by rinse with 10ml cold PBS, followed by aspiration. Thereafter, 5ml of Membrane Scrape Buffer is added to scrape cells; this is followed by transfer of cellular extract into 50ml centrifuge tubes (centrifuged at 20,000 rpm for 17 minutes at 4°C). Thereafter, the supernatant is aspirated and the pellet is resuspended in 30ml Membrane Wash Buffer followed by centrifuge at 20,000 rpm for 17 minutes at 4°C. The  
30 supernatant is then aspirated and the pellet resuspended in Binding Buffer. This is homogenized using a Brinkman Polytron™ homogenizer (15-20 second bursts until the all material is in suspension). This is referred to herein as “Membrane Protein”.

**Bradford Protein Assay**

[0136] Following the homogenization, protein concentration of the membranes is determined using the Bradford Protein Assay (protein can be diluted to about 1.5mg/ml, aliquoted and frozen (-80°C) for later use; when frozen, protocol for use is as follows: on the day of the  
5 assay, frozen Membrane Protein is thawed at room temperature, followed by vortex and then homogenized with a Polytron at about 12 x 1,000 rpm for about 5-10 seconds; it is noted that for multiple preparations, the homogenizer should be thoroughly cleaned between homogenization of different preparations).

**a. Materials**

10 [0137] Binding Buffer (as per above); Bradford Dye Reagent; Bradford Protein Standard will be utilized, following manufacturer instructions (Biorad, cat. no. 500-0006).

**b. Procedure**

[0138] Duplicate tubes are prepared, one including the membrane, and one as a control "blank". Each contained 800µl Binding Buffer. Thereafter, 10µl of Bradford Protein Standard  
15 (1mg/ml) is added to each tube, and 10µl of membrane Protein is then added to just one tube (not the blank). Thereafter, 200µl of Bradford Dye Reagent is added to each tube, followed by vortex of each. After five (5) minutes, the tubes are re-vortexed and the material therein is transferred to cuvettes. The cuvettes are then read using a CECIL 3041 spectrophotometer, at wavelength 595.

**Direct Identification Assay**

20 **a. Materials**

[0139] GDP Buffer consisted of 37.5 ml Binding Buffer and 2mg GDP (Sigma, cat. no. G-7127), followed by a series of dilutions in Binding Buffer to obtain 0.2 µM GDP (final concentration of GDP in each well was 0.1 µM GDP); each well comprising a candidate compound, has a final  
25 volume of 200µl consisting of 100µl GDP Buffer (final concentration, 0.1µM GDP), 50µl Membrane Protein in Binding Buffer, and 50µl [<sup>35</sup>S]GTPγS (0.6 nM) in Binding Buffer (2.5 µl [<sup>35</sup>S]GTPγS per 10ml Binding Buffer).

**b. Procedure**

[0140] Candidate compounds are preferably screened using a 96-well plate format (these  
30 can be frozen at -80°C). Membrane Protein (or membranes with expression vector excluding the GPCR Fusion Protein, as control), is homogenized briefly until in suspension. Protein

concentration is then determined using the Bradford Protein Assay set forth above. Membrane Protein (and control) is diluted to 0.25mg/ml in Binding Buffer (final assay concentration, 12.5µg/well). Thereafter, 100 µl GDP Buffer is added to each well of a Wallac Scintistrip™ (Wallac). A 5ul pin-tool is then used to transfer 5 µl of a candidate compound into such well (i.e., 5µl in total assay volume of 200 µl is a 1:40 ratio such that the final screening concentration of the candidate compound is 10µM). Again, to avoid contamination, after each transfer step the pin tool should be rinsed in three reservoirs comprising water (1X), ethanol (1X) and water (2X) – excess liquid should be shaken from the tool after each rinse and dried with paper and kimwipes. Thereafter, 50 µl of Membrane Protein is added to each well (a control well comprising membranes without the GPCR Fusion Protein was also utilized), and pre-incubated for 5-10 minutes at room temperature. Thereafter, 50µl of [<sup>35</sup>S]GTPγS (0.6 nM) in Binding Buffer is added to each well, followed by incubation on a shaker for 60 minutes at room temperature (again, in this example, plates were covered with foil). The assay is then stopped by spinning of the plates at 4000 RPM for 15 minutes at 22°C. The plates are then aspirated with an 8 channel manifold and sealed with plate covers. The plates are read on a Wallac 1450 using setting “Prot. #37” (as per manufacturer instructions).

### C. Cyclic AMP Assay

[0141] Another assay approach to directly identified candidate compound was accomplished by utilizing a cyclase-based assay. In addition to direct identification, this assay approach can be utilized as an independent approach to provide confirmation of the results from the [<sup>35</sup>S]GTPγS approach as set forth above.

A modified Flash Plate™ Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) is preferably utilized for direct identification of candidate compounds as inverse agonists and agonists to constitutively activated GPCRs in accordance with the following protocol.

[0142] Transfected cells are harvested approximately three days after transfection. Membranes are prepared by homogenization of suspended cells in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl<sub>2</sub>. Homogenization is performed on ice using a Brinkman Polytron™ for approximately 10 seconds. The resulting homogenate is centrifuged at 49,000 X g for 15 minutes at 4°C. The resulting pellet is then resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 x g for 15 minutes at 4°C. The resulting pellet is then stored at -80°C until utilized. On

the day of direct identification screening, the membrane pellet is slowly thawed at room temperature, resuspended in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl<sub>2</sub>, to yield a final protein concentration of 0.60mg/ml (the resuspended membranes are placed on ice until use).

5 [0143] cAMP standards and Detection Buffer (comprising 2  $\mu$ Ci of tracer [<sup>125</sup>I] cAMP (100  $\mu$ l] to 11 ml Detection Buffer) is prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer is prepared fresh for screening and contained 20mM HEPES, pH 7.4, 10mM MgCl<sub>2</sub>, 20mM phosphocreatine (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50  $\mu$ M GTP (Sigma), and 0.2 mM ATP (Sigma); Assay Buffer is then  
10 stored on ice until utilized.

[0144] Candidate compounds identified as per above (if frozen, thawed at room temperature) are added, preferably, to 96-well plate wells (3 $\mu$ l/well; 12 $\mu$ M final assay concentration), together with 40  $\mu$ l Membrane Protein (30 $\mu$ g/well) and 50 $\mu$ l of Assay Buffer. This admixture is then incubated for 30 minutes at room temperature, with gentle shaking.

15 [0145] Following the incubation, 100 $\mu$ l of Detection Buffer is added to each well, followed by incubation for 2-24 hours. Plates are then counted in a Wallac MicroBeta™ plate reader using "Prot. #31" (as per manufacturer instructions).

#### Example 8

#### 20 MELANOPHORE TECHNOLOGY

[0146] Melanophores are skin cells found in lower vertebrates. They contain pigmented organelles termed melanosomes. Melanophores are able to redistribute these melanosomes along a microtubule network upon G-protein coupled receptor (GPCR) activation. The result of this  
25 pigment movement is an apparent lightening or darkening of the cells. In melanophores, the decreased levels of intracellular cAMP that result from activation of a G<sub>i</sub>-coupled receptor cause melanosomes to migrate to the center of the cell, resulting in a dramatic lightening in color. If cAMP levels are then raised, following activation of a G<sub>s</sub>-coupled receptor, the melanosomes are re-dispersed and the cells appear dark again. The increased levels of diacylglycerol that result  
30 from activation of G<sub>q</sub>-coupled receptors can also induce this re-dispersion. In addition, the technology is also suited to the study of certain receptor tyrosine kinases. The response of the melanophores takes place within minutes of receptor activation and results in a simple, robust color change. The response can be easily detected using a conventional absorbance microplate reader or a modest video imaging system. Unlike other skin cells, the melanophores derive from

the neural crest and appear to express a full complement of signaling proteins. In particular, the cells express an extremely wide range of G-proteins and so are able to functionally express almost all GPCRs.

[0147] Melanophores can be utilized to identify compounds, including natural ligands, against GPCRs. This method can be conducted by introducing test cells of a pigment cell line capable of dispersing or aggregating their pigment in response to a specific stimulus and expressing an exogenous clone coding for the GPCR. A stimulant, *e.g.*, melatonin, sets an initial state of pigment disposition wherein the pigment is aggregated within the test cells if activation of the GPCR induces pigment dispersion. However, stimulating the cell with a stimulant to set an initial state of pigment disposition wherein the pigment is dispersed if activation of the GPCR induces pigment aggregation. The test cells are then contacted with chemical compounds, and it is determined whether the pigment disposition in the cells changed from the initial state of pigment disposition. Dispersion of pigments cells due to the candidate compound, including but not limited to a ligand, coupling to the GPCR will appear dark on a petri dish, while aggregation of pigments cells will appear light.

[0148] Materials and methods will be followed according to the disclosure of U.S. Patent Number 5,462,856 and U.S. Patent Number 6,051,386. These patent references are hereby incorporated in their entirety.

[0149] Melanophores are transfected by electroporation with plasmids coding for the GPCRs. The cells are plated in 96-well plates (one receptor per plate). 48 hours post-transfection, half of the cells on each plate are treated with 10nM melatonin. Melatonin activates an endogenous Gi-coupled receptor in the melanophores and causes them to aggregate their pigment. The remaining half of the cells are transferred to serum-free medium 0.7X L-15 (Gibco). After one hour, the cells in serum-free media remain in a pigment-dispersed state while the melatonin-treated cells are in a pigment-aggregated state. At this point, the cells are treated with a dose response of a candidate compound (Sigma). If the plated GPCRs bind to a candidate compound, the melanophores would be expected to undergo a color change in response to the compound. If the receptor is either a G<sub>s</sub> or G<sub>q</sub> coupled receptor, then the melatonin-aggregated melanophores will undergo pigment dispersion. In contrast, if the receptor is a G<sub>i</sub>-coupled receptor, then the pigment-dispersed cells is expected to undergo a dose-dependent pigment aggregation.

[0150] To reconfirm these results, melanophores are transfected with a range of FSHR DNA from 0 to 10 $\mu$ g. As controls, melanophores are also transfected with 10 $\mu$ g of  $\alpha_{2A}$  Adrenergic receptor (a known Gi-coupled receptor) and salmon sperm DNA (Gibco), as a mock

transfection. On day 3, the cells are again incubated for 1 hour in serum-free L-15 medium (Gibco) and remain in a pigment-dispersed state. The cells are then treated with a dose response of the candidate compound.

5 [0151] All references cited throughout this patent document, including co-pending and related patent applications are incorporated herein by reference in their entirety. Modifications and extension of the disclosed inventions that are within the purview of the skilled artisan are encompassed within the above disclosure and the claims that follow.

10 [0152] Although a variety of expression vectors are available to those in the art, for purposes of utilization for both the endogenous and non-endogenous human FSHR, it is most preferred that the vector utilized be pCMV. This vector was deposited with the American Type Culture Collection (ATCC) on October 13, 1998 (10801 University Blvd., Manassas, VA 20110-2209 USA) under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure. The DNA was tested by the ATCC and determined to be viable. The ATCC has assigned the following deposit number to  
15 pCMV: ATCC #203351.

**CLAIMS**

What is claimed is:

- 1 1. An isolated polynucleotide encoding a non-endogenous, constitutively activated version of a  
2 human G protein-coupled receptor, wherein said polynucleotide is selected from the group consisting  
3 of:
  - 4 (a) a polynucleotide comprising a nucleotide sequence encoding the polypeptide of  
5 SEQ ID NO:26; and
  - 6 (b) a polynucleotide comprising the nucleotide sequence of SEQ ID NO:25.
- 1 2. An isolated polynucleotide encoding a non-endogenous, constitutively activated version of a  
2 human G protein-coupled receptor, wherein said polynucleotide is selected from the group consisting  
3 of:
  - 4 (a) a polynucleotide consisting essentially of a nucleotide sequence encoding the  
5 polypeptide of SEQ ID NO:26; and
  - 6 (b) a polynucleotide consisting essentially of the nucleotide sequence of SEQ ID  
7 NO:25.
- 1 3. A vector comprising the polynucleotide of claim 1 or claim 2.
- 1 4. The vector of claim 3, wherein said vector is an expression vector, and said polynucleotide is  
2 operably linked to a promoter.
- 1 5. A recombinant host cell comprising the vector of claim 3.

- 1 6. An isolated polynucleotide encoding a non-endogenous, constitutively activated version of a  
2 human G protein-coupled receptor, wherein said polynucleotide is selected from the group consisting  
3 of:
- 4 (a) a polynucleotide comprising a nucleotide sequence encoding the polypeptide of  
5 SEQ.ID.NO.: 26 wherein the codon corresponding to leucine at amino acid position  
6 460 has been substituted with a codon corresponding to an amino acid other than  
7 leucine; and
- 8 (b) a polynucleotide comprising the nucleotide sequence of SEQ.ID.NO.: 25 wherein  
9 the codon at nucleotide positions 1378-1380 corresponding to leucine has been  
10 substituted with a codon corresponding to an amino acid other than leucine.
- 1 7. An isolated polynucleotide encoding a non-endogenous, constitutively activated G protein-  
2 coupled receptor, wherein said polynucleotide is selected from the group consisting of:
- 3 (a) a polynucleotide consisting essentially of a nucleotide sequence encoding the  
4 polypeptide of SEQ.ID.NO.:26 wherein the codon corresponding to glycine at  
5 amino acid position leucine has been substituted with a codon corresponding to an  
6 amino acid other than leucine; and
- 7 (b) a polynucleotide consisting essentially of the nucleotide sequence of SEQ.ID.NO.:  
8 25 wherein the codon at nucleotide positions 1378-1380 corresponding to leucine  
9 has been substituted with a codon corresponding to an amino acid other than  
10 leucine.
- 1 8. The polynucleotide of claim 6 and 7 wherein the codon corresponding to leucine at amino  
2 acid position 460 or the codon at nucleotide positions 1378-1380 corresponding to leucine has been  
3 substituted with a codon corresponding to arginine.
- 1 9. A vector comprising the polynucleotide of any one of claims 6, 7, or 8.

1 10. A vector of claim 9, wherein said vector is an expression vector, and said polynucleotide is  
2 operably linked to a promoter.

1 11. A recombinant host cell comprising the vector of claim 10.

1 12. A process for making recombinant host cell comprising the steps of:

- 2 (a) transfecting the expression vector of claim 9 into a suitable host cell; and  
3 (b) culturing the host cell under conditions which allow expression of a G protein  
4 fusion construct of a G protein-coupled receptor from the expression vector.

1 13. A membrane of the recombinant host cell of claim 11 comprising said expressed G protein  
2 fusion construct.

Figure 1

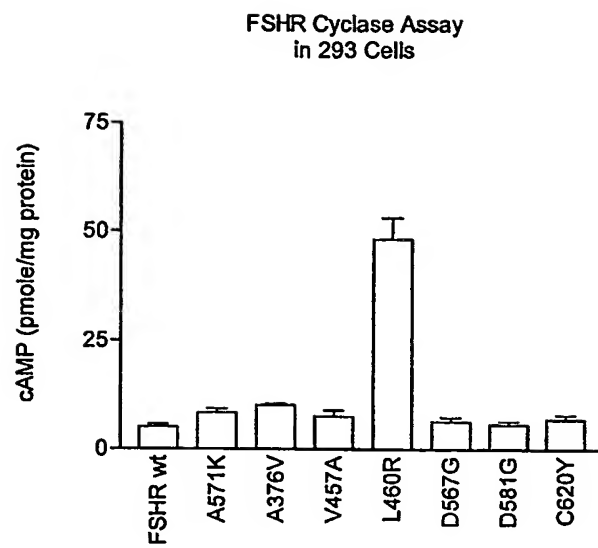


Figure 2

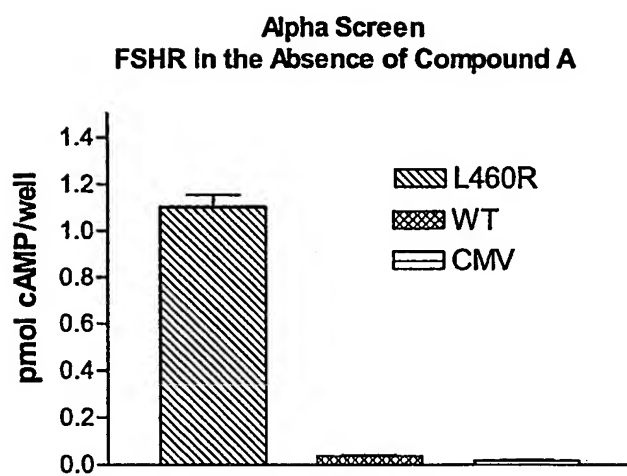
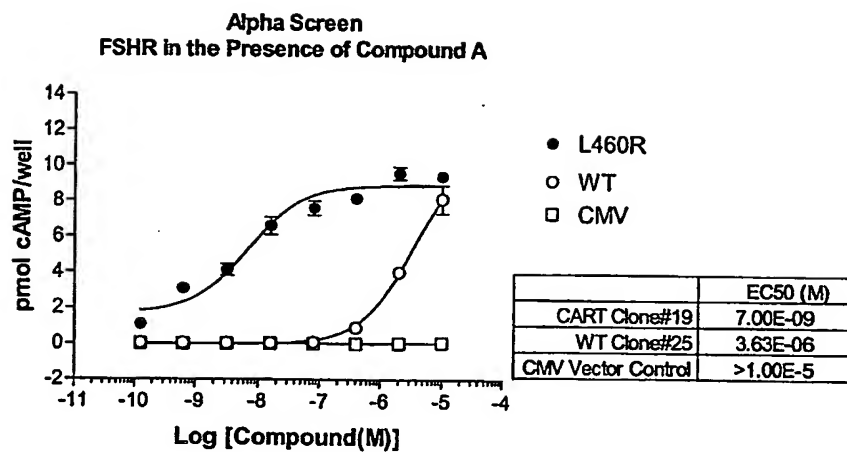


Figure 3



## SEQUENCE LISTING

<110> Arena Pharmaceuticals, Inc.  
Chen W. Liaw

<120> Non-Endogenous Versions of Human G Protein-Coupled Receptor: FSHR

<130> 23.W02

<160> 36

<170> PatentIn version 3.1

<210> 1

<211> 2088

<212> DNA

<213> Homo sapien

<400> 1	
atggccctgc tcctgggtctc ttgctggca ttctgagct tgggctcagg atgtcatcat	60
cggatctgtc actgctctaa cagggttttt ctctgccaa agagcaagg gacagagatt	120
ccttctgacc tcccaggaa tgccattgaa ctgaggtttg tcctcacaa gcttcgagtc	180
atccaaaaag gtgcattttc aggatttggg gacctggaga aaatagagat ctctcagaat	240
gatgtcttgg aggtgataga ggcagatgtg ttctccaacc ttcccaaatt acatgaaatt	300
agaattgaaa aggccaacaa cctgctctac atcacccctg aggccttcca gaaccttccc	360
aaccttcaat atctgttaat atccaacaca ggtattaagc accttccaga tggtcacaag	420
attcattctc tccaaaagg tttacttgac attcaagata acataaacat ccacacaatt	480
gaaagaaatt ctttcgtggg gctgagcttt gaaagtgtga ttctatggct gaataagaat	540
gggattcaag aaatacacia ctgtgcattc aatggaaccc aactagatgc agtgaatcta	600
agcgataata ataatttaga agaattgcct aatgatgttt tccacggagc ctctggacca	660
gtcattctag atatttcaag aacaaggatc cattccctgc ctagctatgg cttagaaaa	720
cttaagaagc tgagggccag gtcgacttac aacttaaaaa agctgcctac tctggaaaa	780

```

cttgtcgccc tcatggaagc cagcctcacc tatcccagcc attgctgtgc ctttgcaaac   840
tggagacggc aaatctctga gcttcaccca atttgcaaca aatctatattt aaggcaagaa   900
gttgattata tgactcaggc taggggtcag agatcctctc tggcagaaga caatgagtcc   960
agctacagca gaggatttga catgacgtac actgagtttg actatgactt atgcaatgaa  1020
gtggttgacg tgacctgctc ccctaagcca gatgcattca acccatgtga agatatcatg  1080
gggtacaaca tcctcagagt cctgatatgg tttatcagca tcctggccat cactgggaac  1140
atcatagtgc tagtgatcct aactaccagc caatataaac tcacagtccc caggttcctt  1200
atgtgcaacc tggcctttgc tgatctctgc attggaatct acctgctgct cattgcatca  1260
gttgatatcc ataccaagag ccaatatcac aactatgcca ttgactggca aactggggca  1320
ggctgtgatg ctgctggctt tttcactgct tttgccagtg agctgtcagt ctacactctg  1380
acagctatca ccttggaag atggcatacc atcacgcatg ccatgcagct ggactgcaag  1440
gtgcagctcc gccatgctgc cagtgtcatg gtgatgggct ggatttttgc ttttgcagct  1500
gccctctttc ccatctttgg catcagcagc tacatgaagg tgagcatctg cctgcccag  1560
gatattgaca gccctttgtc acagctgtat gtcagtgtccc tccttgtgct caatgtcctg  1620
gcctttgtgg tcactgtgg ctgctatata cacatctacc tcacagtgcg gaaccccaac  1680
atcgtgtcct cctctagtga caccaggatc gccaaagcga tggccatgct catcttcact  1740
gacttcctct gcatggcacc catttctttc tttgccattt ctgectccct caagggtgcc  1800
ctcatcactg tgtccaaagc aaagattctg ctggttctgt ttcaccccat caactcctgt  1860
gccaaacctt tcctctatgc catctttacc aaaaactttc gcagagattt cttcattctg  1920
ctgagcaagt gtggctgcta tgaaatgcaa gcccaaattt ataggacaga aacttcatcc  1980
actgtccaca acacccatcc aaggaatggc cactgtctct cagctcccag agtcaccagt  2040
ggttccactt acatacttgt ccctctaagt catttagccc aaaactaa                2088

```

&lt;210&gt; 2

&lt;211&gt; 695

&lt;212&gt; PRT

&lt;213&gt; Homo sapien

&lt;400&gt; 2

```

Met Ala Leu Leu Val Ser Leu Leu Ala Phe Leu Ser Leu Gly Ser
1           5           10          15

```

Gly Cys His His Arg Ile Cys His Cys Ser Asn Arg Val Phe Leu Cys  
 20 25 30

Gln Glu Ser Lys Val Thr Glu Ile Pro Ser Asp Leu Pro Arg Asn Ala  
 35 40 45

Ile Glu Leu Arg Phe Val Leu Thr Lys Leu Arg Val Ile Gln Lys Gly  
 50 55 60

Ala Phe Ser Gly Phe Gly Asp Leu Glu Lys Ile Glu Ile Ser Gln Asn  
 65 70 75 80

Asp Val Leu Glu Val Ile Glu Ala Asp Val Phe Ser Asn Leu Pro Lys  
 85 90 95

Leu His Glu Ile Arg Ile Glu Lys Ala Asn Asn Leu Leu Tyr Ile Thr  
 100 105 110

Pro Glu Ala Phe Gln Asn Leu Pro Asn Leu Gln Tyr Leu Leu Ile Ser  
 115 120 125

Asn Thr Gly Ile Lys His Leu Pro Asp Val His Lys Ile His Ser Leu  
 130 135 140

Gln Lys Val Leu Leu Asp Ile Gln Asp Asn Ile Asn Ile His Thr Ile  
 145 150 155 160

Glu Arg Asn Ser Phe Val Gly Leu Ser Phe Glu Ser Val Ile Leu Trp  
 165 170 175

Leu Asn Lys Asn Gly Ile Gln Glu Ile His Asn Cys Ala Phe Asn Gly  
 180 185 190

Thr Gln Leu Asp Ala Val Asn Leu Ser Asp Asn Asn Asn Leu Glu Glu  
 195 200 205

Leu Pro Asn Asp Val Phe His Gly Ala Ser Gly Pro Val Ile Leu Asp  
 210 215 220

Ile Ser Arg Thr Arg Ile His Ser Leu Pro Ser Tyr Gly Leu Glu Asn  
 225 230 235 240

Leu Lys Lys Leu Arg Ala Arg Ser Thr Tyr Asn Leu Lys Lys Leu Pro  
 245 250 255

Thr Leu Glu Lys Leu Val Ala Leu Met Glu Ala Ser Leu Thr Tyr Pro

260	265	270
Ser His Cys Cys Ala Phe Ala Asn Trp Arg Arg Gln Ile Ser Glu Leu 275 280 285		
His Pro Ile Cys Asn Lys Ser Ile Leu Arg Gln Glu Val Asp Tyr Met 290 295 300		
Thr Gln Ala Arg Gly Gln Arg Ser Ser Leu Ala Glu Asp Asn Glu Ser 305 310 315 320		
Ser Tyr Ser Arg Gly Phe Asp Met Thr Tyr Thr Glu Phe Asp Tyr Asp 325 330 335		
Leu Cys Asn Glu Val Val Asp Val Thr Cys Ser Pro Lys Pro Asp Ala 340 345 350		
Phe Asn Pro Cys Glu Asp Ile Met Gly Tyr Asn Ile Leu Arg Val Leu 355 360 365		
Ile Trp Phe Ile Ser Ile Leu Ala Ile Thr Gly Asn Ile Ile Val Leu 370 375 380		
Val Ile Leu Thr Thr Ser Gln Tyr Lys Leu Thr Val Pro Arg Phe Leu 385 390 395 400		
Met Cys Asn Leu Ala Phe Ala Asp Leu Cys Ile Gly Ile Tyr Leu Leu 405 410 415		
Leu Ile Ala Ser Val Asp Ile His Thr Lys Ser Gln Tyr His Asn Tyr 420 425 430		
Ala Ile Asp Trp Gln Thr Gly Ala Gly Cys Asp Ala Ala Gly Phe Phe 435 440 445		
Thr Val Phe Ala Ser Glu Leu Ser Val Tyr Thr Leu Thr Ala Ile Thr 450 455 460		
Leu Glu Arg Trp His Thr Ile Thr His Ala Met Gln Leu Asp Cys Lys 465 470 475 480		
Val Gln Leu Arg His Ala Ala Ser Val Met Val Met Gly Trp Ile Phe 485 490 495		
Ala Phe Ala Ala Ala Leu Phe Pro Ile Phe Gly Ile Ser Ser Tyr Met 500 505 510		

Lys Val Ser Ile Cys Leu Pro Met Asp Ile Asp Ser Pro Leu Ser Gln  
 515 520 525

Leu Tyr Val Met Ser Leu Leu Val Leu Asn Val Leu Ala Phe Val Val  
 530 535 540

Ile Cys Gly Cys Tyr Ile His Ile Tyr Leu Thr Val Arg Asn Pro Asn  
 545 550 555 560

Ile Val Ser Ser Ser Ser Asp Thr Arg Ile Ala Lys Arg Met Ala Met  
 565 570 575

Leu Ile Phe Thr Asp Phe Leu Cys Met Ala Pro Ile Ser Phe Phe Ala  
 580 585 590

Ile Ser Ala Ser Leu Lys Val Pro Leu Ile Thr Val Ser Lys Ala Lys  
 595 600 605

Ile Leu Leu Val Leu Phe His Pro Ile Asn Ser Cys Ala Asn Pro Phe  
 610 615 620

Leu Tyr Ala Ile Phe Thr Lys Asn Phe Arg Arg Asp Phe Phe Ile Leu  
 625 630 635 640

Leu Ser Lys Cys Gly Cys Tyr Glu Met Gln Ala Gln Ile Tyr Arg Thr  
 645 650 655

Glu Thr Ser Ser Thr Val His Asn Thr His Pro Arg Asn Gly His Cys  
 660 665 670

Ser Ser Ala Pro Arg Val Thr Ser Gly Ser Thr Tyr Ile Leu Val Pro  
 675 680 685

Leu Ser His Leu Ala Gln Asn  
 690 695

<210> 3

<211> 30

<212> DNA

<213> Artificial

<400> 3

atcaccatgg ccctgctcct ggtctctttg

30

&lt;210&gt; 4

&lt;211&gt; 30

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 4

tgccttaaaa tagatttggt gcaaattgga

30

&lt;210&gt; 5

&lt;211&gt; 30

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 5

ctctgagctt catccaattt gcaacaaatc

30

&lt;210&gt; 6

&lt;211&gt; 36

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 6

tgtgaattcg ttttgggcta aatgacttag agggac

36

&lt;210&gt; 7

&lt;211&gt; 31

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 7

ttatcagcat cctggtcac actgggaaca t

31

&lt;210&gt; 8

<211> 31

<212> DNA

<213> Artificial

<400> 8

atgttcccag tgatgaccag gatgctgata a

31

<210> 9

<211> 31

<212> DNA

<213> Artificial

<400> 9

ccagtgaagct gtcagcctac actctgacag c

31

<210> 10

<211> 31

<212> DNA

<213> Artificial

<400> 10

gctgtcagag tgtaggctga cagctcactg g

31

<210> 11

<211> 31

<212> DNA

<213> Artificial

<400> 11

tgtcagtcta cactcggaca gctatcacct t

31

<210> 12

<211> 31

<212> DNA

<213> Artificial

<400> 12

aaggatgatag ctgtccgagt gtagactgac a

31

<210> 13

<211> 31

<212> DNA

<213> Artificial

<400> 13

tgtcctctctc tagtggcacc aggatcgcca a

31

<210> 14

<211> 31

<212> DNA

<213> Artificial

<400> 14

ttggcgatcc tggtgccact agaggaggac a

31

<210> 15

<211> 33

<212> DNA

<213> Artificial

<400> 15

agtgacacca ggatcaagaa gcgcatggcc atg

33

<210> 16

<211> 33

<212> DNA

<213> Artificial

<400> 16

catggccatg cgcttcttga tcctgggtgc act

33

<210> 17

<211> 31

<212> DNA

<213> Artificial

<400> 17

tgctcatctt cactggcttc ctctgcatgg c

31

<210> 18

<211> 31

<212> DNA

<213> Artificial

<400> 18

gccatgcaga ggaagccagt gaagatgagc a

31

<210> 19

<211> 31

<212> DNA

<213> Artificial

<400> 19

accccatcaa ctctatgcc aacccttcc t

31

<210> 20

<211> 31

<212> DNA

<213> Artificial

<400> 20

aggaaggggt tggcatagga gttgatggg t

31

<210> 21

&lt;211&gt; 2088

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 21

atggccctgc tcctggtctc ttgctggca ttctgagct tgggctcagg atgtcatcat	60
cggtactgtc actgctctaa cagggttttt ctctgccaa agagcaagg gacagagatt	120
ccttctgacc tcccaggaa tgccattgaa ctgaggtttg tcctcacaa gcttcagtc	180
atccaaaaag gtgcattttc aggatttggg gacctggaga aaatagagat ctctcagaat	240
gatgtcttgg aggtgataga ggcagatgtg ttctccaacc ttcccaaatt acatgaaatt	300
agaattgaaa aggccaacaa cctgctctac atcaccctg aggccttcca gaaccttccc	360
aaccttcaat atctgttaat atccaacaca ggtattaagc accttcaga tgttcacaag	420
attcattctc tccaaaagg tttacttgac attcaagata acataaacat ccacacaatt	480
gaaagaaatt ctttcgtggg gctgagcttt gaaagtgtga ttctatggct gaataagaat	540
gggattcaag aaatacacia ctgtgcattc aatggaaccc aactagatgc agtgaatcta	600
agcgataata ataatttaga agaattgcct aatgatgttt tccacggagc ctctggacca	660
gtcattctag atatttcaag aacaaggatc cattccctgc ctagctatgg cttagaaaat	720
cttaagaagc tgagggccag gtcgacttac aacttaaaaa agctgcctac tctggaaaag	780
cttgctgccc tcatggaagc cagcctcacc tatcccagcc attgctgtgc ctttgcaaac	840
tggagacggc aaatctctga gcttcatcca atttgcaaca aatctatttt aaggcaagaa	900
gttgattata tgactcaggc taggggtcag agatcctctc tggcagaaga caatgagtcc	960
agctacagca gaggatttga catgacgtac actgagtttg actatgactt atgcaatgaa	1020
gtgggtgacg tgacctgctc ccctaagcca gatgcattca acccatgtga agatatcatg	1080
gggtacaaca tctcagagt cctgatatgg ttatcagca tcctggatcat cactgggaac	1140
atcatagtgc tagtgatcct aactaccagc caatataaac tcacagtccc caggttcctt	1200
atgtgcaacc tggcctttgc tgatctctgc attggaatct acctgctgct cattgcatca	1260
gttgatatcc ataccaagag ccaatatcac aactatgcca ttgactggca aactggggca	1320
ggctgtgatg ctgctggctt ttactctgc ttgcccagtg agctgtcagt ctacactctg	1380
acagctatca ccttggaag atggcatacc atcacgcatg ccatgcagct ggactgcaag	1440
gtgcagctcc gccatgctgc cagtgtcatg gtgatgggct ggatttttgc ttttgagct	1500
gccctctttc ccatctttgg catcagcagc tacatgaagg tgagcatctg cctgcccag	1560
gatattgaca gccctttgtc acagctgtat gtcagtgtcc tccttggtgt caatgtcctg	1620

```

gcctttgtgg tcatctgtgg ctgctatata cacatctacc tcacagtgcg gaacccaac 1680
atcgtgtcct cctctagtga caccaggatc gccaagcgca tggccatgct catcttcact 1740
gacttcctct gcatggcacc catttcttct tttgccattt ctgcctccct caaggtgccc 1800
ctcatcactg tgtccaaagc aaagattctg ctggttctgt ttcaccccat caactcctgt 1860
gccaaccctc tcctctatgc catctttacc aaaaactttc gcagagattt cttcattctg 1920
ctgagcaagt gtggctgcta tgaaatgcaa gcccaaattt ataggacaga aacttcatcc 1980
actgtccaca acacccatcc aaggaatggc cactgctctt cagctcccag agtcaccagt 2040
ggttccactt acatacttgt ccctctaagt catttagccc aaaactaa 2088

```

&lt;210&gt; 22

&lt;211&gt; 695

&lt;212&gt; PRT

&lt;213&gt; Artificial

&lt;400&gt; 22

```

Met Ala Leu Leu Leu Val Ser Leu Leu Ala Phe Leu Ser Leu Gly Ser
1           5           10           15

```

```

Gly Cys His His Arg Ile Cys His Cys Ser Asn Arg Val Phe Leu Cys
          20           25           30

```

```

Gln Glu Ser Lys Val Thr Glu Ile Pro Ser Asp Leu Pro Arg Asn Ala
          35           40           45

```

```

Ile Glu Leu Arg Phe Val Leu Thr Lys Leu Arg Val Ile Gln Lys Gly
          50           55           60

```

```

Ala Phe Ser Gly Phe Gly Asp Leu Glu Lys Ile Glu Ile Ser Gln Asn
65           70           75           80

```

```

Asp Val Leu Glu Val Ile Glu Ala Asp Val Phe Ser Asn Leu Pro Lys
          85           90           95

```

```

Leu His Glu Ile Arg Ile Glu Lys Ala Asn Asn Leu Leu Tyr Ile Thr
          100          105          110

```

```

Pro Glu Ala Phe Gln Asn Leu Pro Asn Leu Gln Tyr Leu Leu Ile Ser
          115          120          125

```

Asn Thr Gly Ile Lys His Leu Pro Asp Val His Lys Ile His Ser Leu  
 130 135 140

Gln Lys Val Leu Leu Asp Ile Gln Asp Asn Ile Asn Ile His Thr Ile  
 145 150 155 160

Glu Arg Asn Ser Phe Val Gly Leu Ser Phe Glu Ser Val Ile Leu Trp  
 165 170 175

Leu Asn Lys Asn Gly Ile Gln Glu Ile His Asn Cys Ala Phe Asn Gly  
 180 185 190

Thr Gln Leu Asp Ala Val Asn Leu Ser Asp Asn Asn Asn Leu Glu Glu  
 195 200 205

Leu Pro Asn Asp Val Phe His Gly Ala Ser Gly Pro Val Ile Leu Asp  
 210 215 220

Ile Ser Arg Thr Arg Ile His Ser Leu Pro Ser Tyr Gly Leu Glu Asn  
 225 230 235 240

Leu Lys Lys Leu Arg Ala Arg Ser Thr Tyr Asn Leu Lys Lys Leu Pro  
 245 250 255

Thr Leu Glu Lys Leu Val Ala Leu Met Glu Ala Ser Leu Thr Tyr Pro  
 260 265 270

Ser His Cys Cys Ala Phe Ala Asn Trp Arg Arg Gln Ile Ser Glu Leu  
 275 280 285

His Pro Ile Cys Asn Lys Ser Ile Leu Arg Gln Glu Val Asp Tyr Met  
 290 295 300

Thr Gln Ala Arg Gly Gln Arg Ser Ser Leu Ala Glu Asp Asn Glu Ser  
 305 310 315 320

Ser Tyr Ser Arg Gly Phe Asp Met Thr Tyr Thr Glu Phe Asp Tyr Asp  
 325 330 335

Leu Cys Asn Glu Val Val Asp Val Thr Cys Ser Pro Lys Pro Asp Ala  
 340 345 350

Phe Asn Pro Cys Glu Asp Ile Met Gly Tyr Asn Ile Leu Arg Val Leu  
 355 360 365

Ile Trp Phe Ile Ser Ile Leu Val Ile Thr Gly Asn Ile Ile Val Leu  
 370 375 380

Val Ile Leu Thr Thr Ser Gln Tyr Lys Leu Thr Val Pro Arg Phe Leu  
 385 390 395 400

Met Cys Asn Leu Ala Phe Ala Asp Leu Cys Ile Gly Ile Tyr Leu Leu  
 405 410 415

Leu Ile Ala Ser Val Asp Ile His Thr Lys Ser Gln Tyr His Asn Tyr  
 420 425 430

Ala Ile Asp Trp Gln Thr Gly Ala Gly Cys Asp Ala Ala Gly Phe Phe  
 435 440 445

Thr Val Phe Ala Ser Glu Leu Ser Val Tyr Thr Leu Thr Ala Ile Thr  
 450 455 460

Leu Glu Arg Trp His Thr Ile Thr His Ala Met Gln Leu Asp Cys Lys  
 465 470 475 480

Val Gln Leu Arg His Ala Ala Ser Val Met Val Met Gly Trp Ile Phe  
 485 490 495

Ala Phe Ala Ala Ala Leu Phe Pro Ile Phe Gly Ile Ser Ser Tyr Met  
 500 505 510

Lys Val Ser Ile Cys Leu Pro Met Asp Ile Asp Ser Pro Leu Ser Gln  
 515 520 525

Leu Tyr Val Met Ser Leu Leu Val Leu Asn Val Leu Ala Phe Val Val  
 530 535 540

Ile Cys Gly Cys Tyr Ile His Ile Tyr Leu Thr Val Arg Asn Pro Asn  
 545 550 555 560

Ile Val Ser Ser Ser Ser Asp Thr Arg Ile Ala Lys Arg Met Ala Met  
 565 570 575

Leu Ile Phe Thr Asp Phe Leu Cys Met Ala Pro Ile Ser Phe Phe Ala  
 580 585 590

Ile Ser Ala Ser Leu Lys Val Pro Leu Ile Thr Val Ser Lys Ala Lys  
 595 600 605

Ile Leu Leu Val Leu Phe His Pro Ile Asn Ser Cys Ala Asn Pro Phe  
 610 615 620

Leu Tyr Ala Ile Phe Thr Lys Asn Phe Arg Arg Asp Phe Phe Ile Leu  
 625 630 635 640

Leu Ser Lys Cys Gly Cys Tyr Glu Met Gln Ala Gln Ile Tyr Arg Thr  
 645 650 655

Glu Thr Ser Ser Thr Val His Asn Thr His Pro Arg Asn Gly His Cys  
 660 665 670

Ser Ser Ala Pro Arg Val Thr Ser Gly Ser Thr Tyr Ile Leu Val Pro  
 675 680 685

Leu Ser His Leu Ala Gln Asn  
 690 695

<210> 23

<211> 2088

<212> DNA

<213> Artificial

<400> 23

atggccctgc tcctggtctc ttgctggca ttctgagct tgggctcagg atgtcatcat	60
cggatctgtc actgctctaa caggggtttt ctctgccaa agagcaagg gacagagatt	120
ccttctgacc tcccaggaa tgccattgaa ctgaggtttg tcctcaccaa gcttcgagtc	180
atccaaaaag gtgcattttc aggatttggg gacctggaga aatagagat ctctcagaat	240
gatgtcttgg aggtgataga ggcagatgtg ttctccaacc ttcccaaatt acatgaaatt	300
agaattgaaa aggccaacaa cctgctctac atcaccctg aggccttcca gaaccttccc	360
aaccttcaat atctgttaat atccaacaca ggtattaagc accttcaga tgttcacaag	420
attcattctc tccaaaagg tttacttgac attcaagata acataaacat ccacacaatt	480
gaaagaaatt ctttctgtgg gctgagcttt gaaagtgtga ttctatggct gaataagaat	540
gggattcaag aaatacacia ctgtgcattc aatggaaccc aactagatgc agtgaatcta	600
agcgataata ataatttaga agaattgcct aatgatgtt tccacggagc ctctggacca	660
gtcattctag atatttcaag aacaaggatc cattccctgc ctgctatgg cttagaaaat	720
cttaagaagc tgagggccag gtcgacttac aacttaaaaa agctgcctac tctggaaaag	780

```

cttgtcgccc tcatggaagc cagcctcacc tatcccagcc attgctgtgc ctttgcaaac   840
tggagacggc aaatctctga gcttcatcca atttgcaaca aatctatttt aaggcaagaa   900
gttgattata tgactcaggc taggggtcag agatcctctc tggcagaaga caatgagtcc   960
agctacagca gaggatttga catgacgtac actgagtttg actatgactt atgcaatgaa  1020
gtggttgacg tgacctgctc ccctaagcca gatgcattca acccatgtga agatatcatg  1080
gggtacaaca tcctcagagt cctgatatgg tttatcagca tcctggccat cactgggaac  1140
atcatagtgc tagtgatcct aactaccagc caatataaac tcacagtccc caggttcctt  1200
atgtgcaacc tggcctttgc tgatctctgc attggaatct acctgctgct cattgcatca  1260
gttgatatcc ataccaagag ccaatatcac aactatgcca ttgactggca aactggggca  1320
ggctgtgatg ctgctggctt tttcactgct tttgccagtg agctgtcagc ctacactctg  1380
acagctatca ccttggaag atggcatacc atcacgcatg ccatgcagct ggactgcaag  1440
gtgcagctcc gccatgctgc cagtgtcatg gtgatgggct ggatttttgc ttttgcaact  1500
gccctctttc ccatcttttg catcagcagc tacatgaagg tgagcatctg cctgcccctg  1560
gatattgaca gccctttgtc acagctgtat gtcatgtccc tccttgctgt caatgtcctg  1620
gcctttgtgg tcatctgtgg ctgctatata cacatctacc tcacagtgcg gaaccccaac  1680
atcgtgtcct cctctagtga caccaggatc gccaaagcga tggccatgct catcttcact  1740
gacttctctc gcatggcacc catttctttc tttgccattt ctgcctccct caagggtgcc  1800
ctcatcactg tgtccaaagc aaagattctg ctggttctgt ttcaccccat caactcctgt  1860
gccaacccct tcctctatgc catctttacc aaaaactttc gcagagattt cttcattctg  1920
ctgagcaagt gtggctgcta tgaaatgcaa gcccaaattt ataggacaga aacttcatcc  1980
actgtccaca acacccatcc aaggaatggc cactgctctt cagctcccag agtcaccagt  2040
ggttccactt acatacttgt cctctaagt catttagccc aaaactaa                2088

```

&lt;210&gt; 24

&lt;211&gt; 695

&lt;212&gt; PRT

&lt;213&gt; Artificial

&lt;400&gt; 24

```

Met Ala Leu Leu Leu Val Ser Leu Leu Ala Phe Leu Ser Leu Gly Ser
1           5           10           15

```

Gly Cys His His Arg Ile Cys His Cys Ser Asn Arg Val Phe Leu Cys  
20 25 30

Gln Glu Ser Lys Val Thr Glu Ile Pro Ser Asp Leu Pro Arg Asn Ala  
35 40 45

Ile Glu Leu Arg Phe Val Leu Thr Lys Leu Arg Val Ile Gln Lys Gly  
50 55 60

Ala Phe Ser Gly Phe Gly Asp Leu Glu Lys Ile Glu Ile Ser Gln Asn  
65 70 75 80

Asp Val Leu Glu Val Ile Glu Ala Asp Val Phe Ser Asn Leu Pro Lys  
85 90 95

Leu His Glu Ile Arg Ile Glu Lys Ala Asn Asn Leu Leu Tyr Ile Thr  
100 105 110

Pro Glu Ala Phe Gln Asn Leu Pro Asn Leu Gln Tyr Leu Leu Ile Ser  
115 120 125

Asn Thr Gly Ile Lys His Leu Pro Asp Val His Lys Ile His Ser Leu  
130 135 140

Gln Lys Val Leu Leu Asp Ile Gln Asp Asn Ile Asn Ile His Thr Ile  
145 150 155 160

Glu Arg Asn Ser Phe Val Gly Leu Ser Phe Glu Ser Val Ile Leu Trp  
165 170 175

Leu Asn Lys Asn Gly Ile Gln Glu Ile His Asn Cys Ala Phe Asn Gly  
180 185 190

Thr Gln Leu Asp Ala Val Asn Leu Ser Asp Asn Asn Asn Leu Glu Glu  
195 200 205

Leu Pro Asn Asp Val Phe His Gly Ala Ser Gly Pro Val Ile Leu Asp  
210 215 220

Ile Ser Arg Thr Arg Ile His Ser Leu Pro Ser Tyr Gly Leu Glu Asn  
225 230 235 240

Leu Lys Lys Leu Arg Ala Arg Ser Thr Tyr Asn Leu Lys Lys Leu Pro  
245 250 255

Thr Leu Glu Lys Leu Val Ala Leu Met Glu Ala Ser Leu Thr Tyr Pro

260	265	270
Ser His Cys Cys Ala Phe Ala Asn Trp Arg Arg Gln Ile Ser Glu Leu		
275	280	285
His Pro Ile Cys Asn Lys Ser Ile Leu Arg Gln Glu Val Asp Tyr Met		
290	295	300
Thr Gln Ala Arg Gly Gln Arg Ser Ser Leu Ala Glu Asp Asn Glu Ser		
305	310	315
Ser Tyr Ser Arg Gly Phe Asp Met Thr Tyr Thr Glu Phe Asp Tyr Asp		
325	330	335
Leu Cys Asn Glu Val Val Asp Val Thr Cys Ser Pro Lys Pro Asp Ala		
340	345	350
Phe Asn Pro Cys Glu Asp Ile Met Gly Tyr Asn Ile Leu Arg Val Leu		
355	360	365
Ile Trp Phe Ile Ser Ile Leu Ala Ile Thr Gly Asn Ile Ile Val Leu		
370	375	380
Val Ile Leu Thr Thr Ser Gln Tyr Lys Leu Thr Val Pro Arg Phe Leu		
385	390	395
Met Cys Asn Leu Ala Phe Ala Asp Leu Cys Ile Gly Ile Tyr Leu Leu		
405	410	415
Leu Ile Ala Ser Val Asp Ile His Thr Lys Ser Gln Tyr His Asn Tyr		
420	425	430
Ala Ile Asp Trp Gln Thr Gly Ala Gly Cys Asp Ala Ala Gly Phe Phe		
435	440	445
Thr Val Phe Ala Ser Glu Leu Ser Ala Tyr Thr Leu Thr Ala Ile Thr		
450	455	460
Leu Glu Arg Trp His Thr Ile Thr His Ala Met Gln Leu Asp Cys Lys		
465	470	475
Val Gln Leu Arg His Ala Ala Ser Val Met Val Met Gly Trp Ile Phe		
485	490	495
Ala Phe Ala Ala Ala Leu Phe Pro Ile Phe Gly Ile Ser Ser Tyr Met		
500	505	510

Lys Val Ser Ile Cys Leu Pro Met Asp Ile Asp Ser Pro Leu Ser Gln  
 515 520 525

Leu Tyr Val Met Ser Leu Leu Val Leu Asn Val Leu Ala Phe Val Val  
 530 535 540

Ile Cys Gly Cys Tyr Ile His Ile Tyr Leu Thr Val Arg Asn Pro Asn  
 545 550 555 560

Ile Val Ser Ser Ser Ser Asp Thr Arg Ile Ala Lys Arg Met Ala Met  
 565 570 575

Leu Ile Phe Thr Asp Phe Leu Cys Met Ala Pro Ile Ser Phe Phe Ala  
 580 585 590

Ile Ser Ala Ser Leu Lys Val Pro Leu Ile Thr Val Ser Lys Ala Lys  
 595 600 605

Ile Leu Leu Val Leu Phe His Pro Ile Asn Ser Cys Ala Asn Pro Phe  
 610 615 620

Leu Tyr Ala Ile Phe Thr Lys Asn Phe Arg Arg Asp Phe Phe Ile Leu  
 625 630 635 640

Leu Ser Lys Cys Gly Cys Tyr Glu Met Gln Ala Gln Ile Tyr Arg Thr  
 645 650 655

Glu Thr Ser Ser Thr Val His Asn Thr His Pro Arg Asn Gly His Cys  
 660 665 670

Ser Ser Ala Pro Arg Val Thr Ser Gly Ser Thr Tyr Ile Leu Val Pro  
 675 680 685

Leu Ser His Leu Ala Gln Asn  
 690 695

<210> 25

<211> 2088

<212> DNA

<213> Artificial

<400> 25

atggccctgc tcctggtctc ttgctggca ttctgagct tgggctcagg atgtcatcat	60
cggatctgtc actgctctaa cagggttttt ctctgccaa agagcaaggt gacagagatt	120
ccttctgacc tcccaggaa tgccattgaa ctgaggtttg tcctcacaa gcttcgagtc	180
atccaaaaag gtgcattttc aggatttggg gacctggaga aaatagagat ctctcagaat	240
gatgtcttgg aggtgataga ggcagatgtg ttctccaacc ttcccaaatt acatgaaatt	300
agaattgaaa aggccaacaa cctgctctac atcacccctg aggccttcca gaaccttccc	360
aaccttcaat atctgttaat atccaacaca ggtattaagc accttccaga tggtcacaag	420
attcattctc tccaaaaggt ttacttgac attcaagata acataaacat ccacacaatt	480
gaaagaaatt ctttctgagg gctgagcttt gaaagtgtga ttctatggct gaataagaat	540
gggattcaag aaatacacia ctgtgcattc aatggaacct aactagatgc agtgaatcta	600
agcgataata ataatttaga agaattgcct aatgatgttt tccacggagc ctctggacca	660
gtcattctag atatttcaag aacaaggatc cattccctgc ctatgatgg cttagaaaat	720
cttaagaagc tgagggccag gtcgacttac aacttaaaaa agctgcctac tctggaaaag	780
cttgctgccc tcatggaagc cagcctcacc tatccagcc attgctgtgc ctttgcaaac	840
tggagacggc aaatctctga gcttcatcca atttgcaaca aatctatttt aaggcaagaa	900
gttgattata tgactcaggc taggggtcag agatcctctc tggcagaaga caatgagtcc	960
agctacagca gaggatttga catgacgtac actgagtttg actatgactt atgcaatgaa	1020
gtggttgacg tgacctgtc ccctaagcca gatgcattca acccatgtga agatatcatg	1080
gggtacaaca tcctcagagt cctgatatgg ttatcagca tcctggccat cactgggaac	1140
atcatagtgc tagtgatcct aactaccagc caatataaac tcacagtccc caggttcctt	1200
atgtgcaacc tggcctttgc tgatctctgc attggaatct acctgctget cattgcatca	1260
gttgatatcc ataccaagag ccaatatcac aactatgcca ttgactggca aactggggca	1320
ggctgtgatg ctgctggcct ttactgtgc ttgcccagtg agctgtcagt ctacactcgg	1380
acagctatca ccttggaag atggcatacc atcacgcatg ccatgcagct ggactgcaag	1440
gtgcagctcc gccatgctgc cagtgtcatg gtgatgggct ggatttttgc ttttgcagct	1500
gccctctttc ccatctttgg catcagcagc tacatgaagg tgagcatctg cctgcccctg	1560
gatattgaca gccctttgtc acagctgtat gtcatgtccc tccttggtgt caatgtcctg	1620
gcctttgtgg tcatctgtgg ctgctatata cacatctacc tcacagtgcg gaacccaac	1680
atcgtgtcct cctctagtga caccaggatc gccaaagcga tggccatgct catcttact	1740
gacttctctc gcatggcacc cttttcttct ttgcccattt ctgctccct caaggtgccc	1800
ctcatcactg tgtccaaagc aaagattctg ctggttctgt ttcaccccat caactcctgt	1860

gccaaccctt tcctctatgc catctttacc aaaaactttc gcagagattt cttcattctg 1920  
 ctgagcaagt gtggctgcta tgaaatgcaa gcccaaattt ataggacaga aacttcatcc 1980  
 actgtccaca acacccatcc aaggaatggc cactgctctt cagctcccag agtcaccagt 2040  
 ggttccactt acatacttgt ccctctaagt catttagccc aaaactaa 2088

<210> 26

<211> 695

<212> PRT

<213> Artificial

<400> 26

Met Ala Leu Leu Val Ser Leu Leu Ala Phe Leu Ser Leu Gly Ser  
 1 5 10 15

Gly Cys His His Arg Ile Cys His Cys Ser Asn Arg Val Phe Leu Cys  
 20 25 30

Gln Glu Ser Lys Val Thr Glu Ile Pro Ser Asp Leu Pro Arg Asn Ala  
 35 40 45

Ile Glu Leu Arg Phe Val Leu Thr Lys Leu Arg Val Ile Gln Lys Gly  
 50 55 60

Ala Phe Ser Gly Phe Gly Asp Leu Glu Lys Ile Glu Ile Ser Gln Asn  
 65 70 75 80

Asp Val Leu Glu Val Ile Glu Ala Asp Val Phe Ser Asn Leu Pro Lys  
 85 90 95

Leu His Glu Ile Arg Ile Glu Lys Ala Asn Asn Leu Leu Tyr Ile Thr  
 100 105 110

Pro Glu Ala Phe Gln Asn Leu Pro Asn Leu Gln Tyr Leu Leu Ile Ser  
 115 120 125

Asn Thr Gly Ile Lys His Leu Pro Asp Val His Lys Ile His Ser Leu  
 130 135 140

Gln Lys Val Leu Leu Asp Ile Gln Asp Asn Ile Asn Ile His Thr Ile  
 145 150 155 160

Glu Arg Asn Ser Phe Val Gly Leu Ser Phe Glu Ser Val Ile Leu Trp  
 165 170 175

Leu Asn Lys Asn Gly Ile Gln Glu Ile His Asn Cys Ala Phe Asn Gly  
 180 185 190

Thr Gln Leu Asp Ala Val Asn Leu Ser Asp Asn Asn Asn Leu Glu Glu  
 195 200 205

Leu Pro Asn Asp Val Phe His Gly Ala Ser Gly Pro Val Ile Leu Asp  
 210 215 220

Ile Ser Arg Thr Arg Ile His Ser Leu Pro Ser Tyr Gly Leu Glu Asn  
 225 230 235 240

Leu Lys Lys Leu Arg Ala Arg Ser Thr Tyr Asn Leu Lys Lys Leu Pro  
 245 250 255

Thr Leu Glu Lys Leu Val Ala Leu Met Glu Ala Ser Leu Thr Tyr Pro  
 260 265 270

Ser His Cys Cys Ala Phe Ala Asn Trp Arg Arg Gln Ile Ser Glu Leu  
 275 280 285

His Pro Ile Cys Asn Lys Ser Ile Leu Arg Gln Glu Val Asp Tyr Met  
 290 295 300

Thr Gln Ala Arg Gly Gln Arg Ser Ser Leu Ala Glu Asp Asn Glu Ser  
 305 310 315 320

Ser Tyr Ser Arg Gly Phe Asp Met Thr Tyr Thr Glu Phe Asp Tyr Asp  
 325 330 335

Leu Cys Asn Glu Val Val Asp Val Thr Cys Ser Pro Lys Pro Asp Ala  
 340 345 350

Phe Asn Pro Cys Glu Asp Ile Met Gly Tyr Asn Ile Leu Arg Val Leu  
 355 360 365

Ile Trp Phe Ile Ser Ile Leu Ala Ile Thr Gly Asn Ile Ile Val Leu  
 370 375 380

Val Ile Leu Thr Thr Ser Gln Tyr Lys Leu Thr Val Pro Arg Phe Leu  
 385 390 395 400

Met Cys Asn Leu Ala Phe Ala Asp Leu Cys Ile Gly Ile Tyr Leu Leu  
 405 410 415

Leu Ile Ala Ser Val Asp Ile His Thr Lys Ser Gln Tyr His Asn Tyr  
 420 425 430

Ala Ile Asp Trp Gln Thr Gly Ala Gly Cys Asp Ala Ala Gly Phe Phe  
 435 440 445

Thr Val Phe Ala Ser Glu Leu Ser Val Tyr Thr Arg Thr Ala Ile Thr  
 450 455 460

Leu Glu Arg Trp His Thr Ile Thr His Ala Met Gln Leu Asp Cys Lys  
 465 470 475 480

Val Gln Leu Arg His Ala Ala Ser Val Met Val Met Gly Trp Ile Phe  
 485 490 495

Ala Phe Ala Ala Ala Leu Phe Pro Ile Phe Gly Ile Ser Ser Tyr Met  
 500 505 510

Lys Val Ser Ile Cys Leu Pro Met Asp Ile Asp Ser Pro Leu Ser Gln  
 515 520 525

Leu Tyr Val Met Ser Leu Leu Val Leu Asn Val Leu Ala Phe Val Val  
 530 535 540

Ile Cys Gly Cys Tyr Ile His Ile Tyr Leu Thr Val Arg Asn Pro Asn  
 545 550 555 560

Ile Val Ser Ser Ser Ser Asp Thr Arg Ile Ala Lys Arg Met Ala Met  
 565 570 575

Leu Ile Phe Thr Asp Phe Leu Cys Met Ala Pro Ile Ser Phe Phe Ala  
 580 585 590

Ile Ser Ala Ser Leu Lys Val Pro Leu Ile Thr Val Ser Lys Ala Lys  
 595 600 605

Ile Leu Leu Val Leu Phe His Pro Ile Asn Ser Cys Ala Asn Pro Phe  
 610 615 620

Leu Tyr Ala Ile Phe Thr Lys Asn Phe Arg Arg Asp Phe Phe Ile Leu  
 625 630 635 640

Leu Ser Lys Cys Gly Cys Tyr Glu Met Gln Ala Gln Ile Tyr Arg Thr

[illegible]

**<210> 27**

<211> 2088

<212> DNA

<213> Artificial

<400> 27	
atggccctgc tctctggtctc tttgctggca ttcctgagct tgggctcagg atgtcatcat	60
cggatctgtc actgctctaa caggggtttt ctctgccaa agagcaagg gacagagatt	120
ccttctgacc tcccgaggaa tgccattgaa ctgaggttg tcttcaccaa gcttcgagtc	180
atccaaaaag gtgcattttc aggatttggg gacctggaga aaatagagat ctctcagaat	240
gatgtcttgg aggtgataga ggcagatgtg ttctccaacc ttcccaaatt acatgaaatt	300
agaattgaaa aggccaaaca cctgctctac atcaccctg aggccttcca gaaccttccc	360
aaccttcaat atctgttaat atccaacaca ggtattaagc accttcaga tgttcacaag	420
attcattctc tccaaaaggt tttacttgac attcaagata acataaacat ccacacaatt	480
gaaagaaatt ctttcgtggg gctgagcttt gaaagtgtga ttctatggct gaataagaat	540
gggattcaag aaatacaca ctgtgcattc aatggaaccc aactagatgc agtgaatcta	600
agcgataata ataatttaga agaattgcct aatgatgttt tccacggagc ctctggacca	660
gtcattctag atatttcaag aacaaggatc cattccctgc cttagctatgg cttagaaaat	720
cttaagaagc tgagggccag gtcgacttac aacttaaaa agctgcctac tctggaaaag	780
cttgctgccc tcatggaagc cagcctcacc tatcccagcc attgctgtgc ctttgcaaac	840
tggagacggc aaatctctga gcttcatcca atttgcaaca aatctatttt aaggcaagaa	900
gttgattata tgactcaggc taggggtcag agatcctctc tggcagaaga caatgagtcc	960
agctacagca gaggatttga catgacgtac actgagtttg actatgactt atgcaatgaa	1020
gtggttgacg tgacctgtc ccctaagcca gatgcattca acccatgtga agatatcatg	1080

```

gggtacaaca tcctcagagt cctgatatgg tttatcagca tcctggccat cactgggaac 1140
atcatagtgc tagtgatcct aactaccagc caatataaac tcacagtccc caggttcctt 1200
atgtgcaacc tggcctttgc tgatctctgc attggaatct acctgctgct cattgcatca 1260
gttgatatcc ataccaagag ccaatatcac aactatgcca ttgactggca aactggggca 1320
ggctgtgatg ctgctggcct tttactgtc tttgccagtg agctgtcagt ctacactctg 1380
acagctatca ccttggaag atggcatacc atcacgcatg ccatgcagct ggactgcaag 1440
gtgcagctcc gccatgctgc cagtgtcatg gtgatgggct ggatttttgc ttttgcagct 1500
gccctctttc ccatctttgg catcagcagc tacatgaagg tgagcatctg cctgcccattg 1560
gatattgaca gccctttgtc acagctgtat gtcattgtccc tccttgtgct caatgtcctg 1620
gcctttgtgg tcattctgtg ctgctatctc cacatctacc tcacagtgcg gaaccccaac 1680
atcgtgtcct cctctagtgg caccaggatc gccaaagcga tggccatgct catcttcact 1740
gaattcctct gcatggcacc catttctttc tttgccattt ctgcctccct caaggtgccc 1800
ctcatcactg tgtccaaagc aaagattctg ctggttctgt ttcaccccat caactcctgt 1860
gccaaaccct tcctctatgc catctttacc aaaaactttc gcagagattt cttcattctg 1920
ctgagcaagt gtggctgcta tgaaatgcaa gcccaaattt ataggacaga aacttcatcc 1980
actgtccaca acacccatcc aaggaatggc cactgtctct cagctcccag agtcaccagt 2040
ggttccactt acatacttgt ccctctaagt catttagccc aaaactaa 2088

```

<210> 28

<211> 695

<212> PRT

<213> Artificial

<400> 28

```

Met Ala Leu Leu Leu Val Ser Leu Leu Ala Phe Leu Ser Leu Gly Ser
1           5           10          15

```

```

Gly Cys His His Arg Ile Cys His Cys Ser Asn Arg Val Phe Leu Cys
          20          25          30

```

```

Gln Glu Ser Lys Val Thr Glu Ile Pro Ser Asp Leu Pro Arg Asn Ala
          35          40          45

```

```

Ile Glu Leu Arg Phe Val Leu Thr Lys Leu Arg Val Ile Gln Lys Gly

```

50	55	60
Ala Phe Ser Gly Phe Gly Asp Leu Glu Lys Ile Glu Ile Ser Gln Asn 65 70 75 80		
Asp Val Leu Glu Val Ile Glu Ala Asp Val Phe Ser Asn Leu Pro Lys 85 90 95		
Leu His Glu Ile Arg Ile Glu Lys Ala Asn Asn Leu Leu Tyr Ile Thr 100 105 110		
Pro Glu Ala Phe Gln Asn Leu Pro Asn Leu Gln Tyr Leu Leu Ile Ser 115 120 125		
Asn Thr Gly Ile Lys His Leu Pro Asp Val His Lys Ile His Ser Leu 130 135 140		
Gln Lys Val Leu Leu Asp Ile Gln Asp Asn Ile Asn Ile His Thr Ile 145 150 155 160		
Glu Arg Asn Ser Phe Val Gly Leu Ser Phe Glu Ser Val Ile Leu Trp 165 170 175		
Leu Asn Lys Asn Gly Ile Gln Glu Ile His Asn Cys Ala Phe Asn Gly 180 185 190		
Thr Gln Leu Asp Ala Val Asn Leu Ser Asp Asn Asn Asn Leu Glu Glu 195 200 205		
Leu Pro Asn Asp Val Phe His Gly Ala Ser Gly Pro Val Ile Leu Asp 210 215 220		
Ile Ser Arg Thr Arg Ile His Ser Leu Pro Ser Tyr Gly Leu Glu Asn 225 230 235 240		
Leu Lys Lys Leu Arg Ala Arg Ser Thr Tyr Asn Leu Lys Lys Leu Pro 245 250 255		
Thr Leu Glu Lys Leu Val Ala Leu Met Glu Ala Ser Leu Thr Tyr Pro 260 265 270		
Ser His Cys Cys Ala Phe Ala Asn Trp Arg Arg Gln Ile Ser Glu Leu 275 280 285		
His Pro Ile Cys Asn Lys Ser Ile Leu Arg Gln Glu Val Asp Tyr Met 290 295 300		

Thr Gln Ala Arg Gly Gln Arg Ser Ser Leu Ala Glu Asp Asn Glu Ser  
305 310 315 320

Ser Tyr Ser Arg Gly Phe Asp Met Thr Tyr Thr Glu Phe Asp Tyr Asp  
325 330 335

Leu Cys Asn Glu Val Val Asp Val Thr Cys Ser Pro Lys Pro Asp Ala  
340 345 350

Phe Asn Pro Cys Glu Asp Ile Met Gly Tyr Asn Ile Leu Arg Val Leu  
355 360 365

Ile Trp Phe Ile Ser Ile Leu Ala Ile Thr Gly Asn Ile Ile Val Leu  
370 375 380

Val Ile Leu Thr Thr Ser Gln Tyr Lys Leu Thr Val Pro Arg Phe Leu  
385 390 395 400

Met Cys Asn Leu Ala Phe Ala Asp Leu Cys Ile Gly Ile Tyr Leu Leu  
405 410 415

Leu Ile Ala Ser Val Asp Ile His Thr Lys Ser Gln Tyr His Asn Tyr  
420 425 430

Ala Ile Asp Trp Gln Thr Gly Ala Gly Cys Asp Ala Ala Gly Phe Phe  
435 440 445

Thr Val Phe Ala Ser Glu Leu Ser Val Tyr Thr Leu Thr Ala Ile Thr  
450 455 460

Leu Glu Arg Trp His Thr Ile Thr His Ala Met Gln Leu Asp Cys Lys  
465 470 475 480

Val Gln Leu Arg His Ala Ala Ser Val Met Val Met Gly Trp Ile Phe  
485 490 495

Ala Phe Ala Ala Ala Leu Phe Pro Ile Phe Gly Ile Ser Ser Tyr Met  
500 505 510

Lys Val Ser Ile Cys Leu Pro Met Asp Ile Asp Ser Pro Leu Ser Gln  
515 520 525

Leu Tyr Val Met Ser Leu Leu Val Leu Asn Val Leu Ala Phe Val Val  
530 535 540

Ile Cys Gly Cys Tyr Ile His Ile Tyr Leu Thr Val Arg Asn Pro Asn  
545 550 555 560

Ile Val Ser Ser Ser Ser Gly Thr Arg Ile Ala Lys Arg Met Ala Met  
565 570 575

Leu Ile Phe Thr Asp Phe Leu Cys Met Ala Pro Ile Ser Phe Phe Ala  
580 585 590

Ile Ser Ala Ser Leu Lys Val Pro Leu Ile Thr Val Ser Lys Ala Lys  
595 600 605

Ile Leu Leu Val Leu Phe His Pro Ile Asn Ser Cys Ala Asn Pro Phe  
610 615 620

Leu Tyr Ala Ile Phe Thr Lys Asn Phe Arg Arg Asp Phe Phe Ile Leu  
625 630 635 640

Leu Ser Lys Cys Gly Cys Tyr Glu Met Gln Ala Gln Ile Tyr Arg Thr  
645 650 655

Glu Thr Ser Ser Thr Val His Asn Thr His Pro Arg Asn Gly His Cys  
660 665 670

Ser Ser Ala Pro Arg Val Thr Ser Gly Ser Thr Tyr Ile Leu Val Pro  
675 680 685

Leu Ser His Leu Ala Gln Asn  
690 695

<210> 29

<211> 2088

<212> DNA

<213> Artificial

<400> 29

atggccctgc tcttgggtctc ttgtctggca ttcttgagct tgggctcagg atgtcatcat	60
cggatctgtc actgctctaa cagggttttt ctctgccaa agagcaagg gacagagatt	120
ccttctgacc tcccgaggaa tgccattgaa ctgaggtttg tcttcaccaa gcttcgagtc	180
atccaaaaag gtgcattttc aggatttggg gacctggaga aaatagagat ctctcagaat	240
gatgtcttgg aggtgataga ggcagatgtg ttctccaacc ttcccaaatt acatgaaatt	300

agaattgaaa aggccaaaca cctgctctac atcacccctg aggccttcca gaaccttccc	360
aaccttcaat atctgttaat atccaacaca ggtattaagc accttccaga tggtcacaag	420
attcattctc tccaaaaggt ttacttgac attcaagata acataaacat ccacacaatt	480
gaaagaaatt ctttcgtggg gctgagcttt gaaagtgtga ttctatggct gaataagaat	540
gggattcaag aaatacacia ctgtgcattc aatggaaccc aactagatgc agtgaatcta	600
agcgataata ataatttaga agaattgcct aatgatgttt tccacggagc ctctggacca	660
gtcattctag atatttcaag aacaaggatc cattccctgc ctagctatgg cttagaaaat	720
cttaagaagc tgagggccag gtcgacttac aacttaaaaa agctgcctac tctggaaaag	780
cttgctgccc tcatggaagc cagcctcacc tatcccagcc attgctgtgc ctttgcaaac	840
tggagacggc aaatctctga gcttcatcca atttgcaaca aatctatttt aaggcaagaa	900
gttgattata tgactcaggc taggggtcag agatcctctc tggcagaaga caatgagtcc	960
agctacagca gaggatttga catgacgtac actgagtttg actatgactt atgcaatgaa	1020
gtggttgacg tgacctgctc ccctaagcca gatgcattca acccatgtga agatatcatg	1080
gggtacaaca tcctcagagt cctgatatgg ttatcagca tcctggccat cactgggaac	1140
atcatagtgc tagtgatcct aactaccagc caatataaac tcacagtccc caggttcctt	1200
atgtgcaacc tggcctttgc tgatctctgc attggaatct acctgctgct cattgcatca	1260
gttgatatcc ataccaagag ccaatatcac aactatgcca ttgactggca aactggggca	1320
ggctgtgatg ctgctggctt ttactgtgc ttggccagt agctgtcagt ctacactctg	1380
acagctatca ccttggaag atggcatacc atcacgcatg ccatgcagct ggactgcaag	1440
gtgcagctcc gccatgctgc cagtgtcatg gtgatgggct ggatttttgc ttttgagct	1500
gccctctttc ccatctttgg catcagcagc tacatgaagg tgagcatctg cctgcccag	1560
gatattgaca gccctttgtc acagctgtat gtcagtgtccc tccttggtgt caatgtcctg	1620
gcctttgtgg tcatctgtgg ctgctatatc cacatctacc tcacagtgcg gaacccaac	1680
atcgtgtcct cctctagtga caccaggatc aagaagcgca tggccatgct catcttcact	1740
gacttcctct gcatggcacc catctcttc tttgccattt ctgcctccct caaggtgccc	1800
ctcatcactg tgtccaaagc aaagattctg ctggttctgt ttcaccccat caactcctgt	1860
gccaaccctc tcctctatgc catctttacc aaaaacttgc gcagagattt cttcattctg	1920
ctgagcaagt gtggctgcta tgaaatgcaa gcccgaattt ataggacaga aacttcatcc	1980
actgtccaca acacccatcc aaggaatggc cactgtctct cagctcccag agtcaccagt	2040
ggttccactt acatacttgt ccctctaagt catttagccc aaaactaa	2088

&lt;210&gt; 30

&lt;211&gt; 695

&lt;212&gt; PRT

&lt;213&gt; Artificial

&lt;400&gt; 30

Met Ala Leu Leu Leu Val Ser Leu Leu Ala Phe Leu Ser Leu Gly Ser  
 1 5 10 15

Gly Cys His His Arg Ile Cys His Cys Ser Asn Arg Val Phe Leu Cys  
 20 25 30

Gln Glu Ser Lys Val Thr Glu Ile Pro Ser Asp Leu Pro Arg Asn Ala  
 35 40 45

Ile Glu Leu Arg Phe Val Leu Thr Lys Leu Arg Val Ile Gln Lys Gly  
 50 55 60

Ala Phe Ser Gly Phe Gly Asp Leu Glu Lys Ile Glu Ile Ser Gln Asn  
 65 70 75 80

Asp Val Leu Glu Val Ile Glu Ala Asp Val Phe Ser Asn Leu Pro Lys  
 85 90 95

Leu His Glu Ile Arg Ile Glu Lys Ala Asn Asn Leu Leu Tyr Ile Thr  
 100 105 110

Pro Glu Ala Phe Gln Asn Leu Pro Asn Leu Gln Tyr Leu Leu Ile Ser  
 115 120 125

Asn Thr Gly Ile Lys His Leu Pro Asp Val His Lys Ile His Ser Leu  
 130 135 140

Gln Lys Val Leu Leu Asp Ile Gln Asp Asn Ile Asn Ile His Thr Ile  
 145 150 155 160

Glu Arg Asn Ser Phe Val Gly Leu Ser Phe Glu Ser Val Ile Leu Trp  
 165 170 175

Leu Asn Lys Asn Gly Ile Gln Glu Ile His Asn Cys Ala Phe Asn Gly  
 180 185 190

Thr Gln Leu Asp Ala Val Asn Leu Ser Asp Asn Asn Asn Leu Glu Glu  
 195 200 205

Leu Pro Asn Asp Val Phe His Gly Ala Ser Gly Pro Val Ile Leu Asp  
 210 215 220

Ile Ser Arg Thr Arg Ile His Ser Leu Pro Ser Tyr Gly Leu Glu Asn  
 225 230 235 240

Leu Lys Lys Leu Arg Ala Arg Ser Thr Tyr Asn Leu Lys Lys Leu Pro  
 245 250 255

Thr Leu Glu Lys Leu Val Ala Leu Met Glu Ala Ser Leu Thr Tyr Pro  
 260 265 270

Ser His Cys Cys Ala Phe Ala Asn Trp Arg Arg Gln Ile Ser Glu Leu  
 275 280 285

His Pro Ile Cys Asn Lys Ser Ile Leu Arg Gln Glu Val Asp Tyr Met  
 290 295 300

Thr Gln Ala Arg Gly Gln Arg Ser Ser Leu Ala Glu Asp Asn Glu Ser  
 305 310 315 320

Ser Tyr Ser Arg Gly Phe Asp Met Thr Tyr Thr Glu Phe Asp Tyr Asp  
 325 330 335

Leu Cys Asn Glu Val Val Asp Val Thr Cys Ser Pro Lys Pro Asp Ala  
 340 345 350

Phe Asn Pro Cys Glu Asp Ile Met Gly Tyr Asn Ile Leu Arg Val Leu  
 355 360 365

Ile Trp Phe Ile Ser Ile Leu Ala Ile Thr Gly Asn Ile Ile Val Leu  
 370 375 380

Val Ile Leu Thr Thr Ser Gln Tyr Lys Leu Thr Val Pro Arg Phe Leu  
 385 390 395 400

Met Cys Asn Leu Ala Phe Ala Asp Leu Cys Ile Gly Ile Tyr Leu Leu  
 405 410 415

Leu Ile Ala Ser Val Asp Ile His Thr Lys Ser Gln Tyr His Asn Tyr  
 420 425 430

Ala Ile Asp Trp Gln Thr Gly Ala Gly Cys Asp Ala Ala Gly Phe Phe

435	440	445
Thr Val Phe Ala Ser Glu Leu Ser Val Tyr Thr Leu Thr Ala Ile Thr		
450	455	460
Leu Glu Arg Trp His Thr Ile Thr His Ala Met Gln Leu Asp Cys Lys		
465	470	475 480
Val Gln Leu Arg His Ala Ala Ser Val Met Val Met Gly Trp Ile Phe		
	485	490 495
Ala Phe Ala Ala Ala Leu Phe Pro Ile Phe Gly Ile Ser Ser Tyr Met		
	500	505 510
Lys Val Ser Ile Cys Leu Pro Met Asp Ile Asp Ser Pro Leu Ser Gln		
	515	520 525
Leu Tyr Val Met Ser Leu Leu Val Leu Asn Val Leu Ala Phe Val Val		
	530	535 540
Ile Cys Gly Cys Tyr Ile His Ile Tyr Leu Thr Val Arg Asn Pro Asn		
545	550	555 560
Ile Val Ser Ser Ser Ser Asp Thr Arg Ile Lys Lys Arg Met Ala Met		
	565	570 575
Leu Ile Phe Thr Asp Phe Leu Cys Met Ala Pro Ile Ser Phe Phe Ala		
	580	585 590
Ile Ser Ala Ser Leu Lys Val Pro Leu Ile Thr Val Ser Lys Ala Lys		
	595	600 605
Ile Leu Leu Val Leu Phe His Pro Ile Asn Ser Cys Ala Asn Pro Phe		
	610	615 620
Leu Tyr Ala Ile Phe Thr Lys Asn Phe Arg Arg Asp Phe Phe Ile Leu		
625	630	635 640
Leu Ser Lys Cys Gly Cys Tyr Glu Met Gln Ala Gln Ile Tyr Arg Thr		
	645	650 655
Glu Thr Ser Ser Thr Val His Asn Thr His Pro Arg Asn Gly His Cys		
	660	665 670
Ser Ser Ala Pro Arg Val Thr Ser Gly Ser Thr Tyr Ile Leu Val Pro		
	675	680 685

Leu Ser His Leu Ala Gln Asn  
690 695

<210> 31

<211> 2088

<212> DNA

<213> Artificial

<400> 31

atggccctgc tcttggtctc ttgctggca ttctgagct tgggctcagg atgtcatcat	60
cggatctgtc actgctctaa caggggtttt ctctgccaa agagcaagg gacagagatt	120
ccttctgacc tcccaggaa tgccattgaa ctgaggttg tctcaccaa gcttcgagtc	180
atccaaaaag gtgcattttc aggatttggg gacctggaga aaatagagat ctctcagaat	240
gatgtcttgg aggtgataga ggcagatgtg ttctccaacc ttcccaaatt acatgaaatt	300
agaattgaaa aggccacaa cctgctctac atcacccctg aggccttcca gaaccttccc	360
aaccttcaat atctgttaat atccaacaca ggtattaagc acctccaga tgttcacaag	420
attcattctc tccaaaagg tttacttgac attcaagata acataaacat ccacacaatt	480
gaaagaaatt ctttcgtggg gctgagcttt gaaagtgtga ttctatggct gaataagaat	540
gggattcaag aaatacacia ctgtgcattc aatggaacc aactagatgc agtgaatcta	600
agcgataata ataatttaga agaattgcct aatgatgttt tccacggagc ctctggacca	660
gtcattctag atatttcaag aacaaggatc cattccctgc ctagctatgg cttagaaaat	720
cttaagaagc tgagggccag gtcgacttac aacttaaaaa agctgcctac tctggaaaag	780
cttgctgccc tcatggaagc cagcctcacc tatcccagcc attgctgtgc ctttgcaaac	840
tggagacggc aaatctctga gcttcatcca atttgcaaca aatctatttt aaggcaagaa	900
gttgattata tgactcaggc taggggtcag agatcctctc tggcagaaga caatgagtcc	960
agctacagca gaggatttga catgacgtac actgagtttg actatgactt atgcaatgaa	1020
gtggttgacg tgacctgctc ccctaagcca gatgcattca acccatgtga agatatcatg	1080
gggtacaaca tctcagagt cctgatatgg ttatcagca tcttgccat cactgggaac	1140
atcatagtgc tagtgatcct aactaccagc caatataaac tcacagtccc caggttcctt	1200
atgtgcaacc tggcctttgc tgatctctgc attggaatct acctgtgct cattgcatca	1260
gttgatatcc ataccaagag ccaatatcac aactatgcca ttgactggca aactggggca	1320

```

ggctgtgatg ctgctggctt tttcactgtc ttggcagtg agctgtcagt ctacactctg 1380
acagctatca ccttggaag atggcatacc atcacgcatg ccatgcagct ggactgcaag 1440
gtgcagctcc gccatgctgc cagtgtcatg gtgatgggct ggatttttgc ttttgcatg 1500
gccctctttc ccatctttgg catcagcagc tacatgaagg tgagcatctg cctgccccatg 1560
gatattgaca gccctttgtc acagctgtat gtcatgtccc tccttgtgct caatgtcctg 1620
gcctttgtgg tcactgtgg ctgctatata cacatctacc tcacagtgcg gaacccaac 1680
atcgtgtcct cctctagtga caccaggatc gccaagcgca tggccatgct catcttcact 1740
ggcttcctct gcatggcacc catttctttc ttggcattt ctgcctccct caagggtccc 1800
ctcatcactg tgtccaaagc aaagattctg ctggttctgt ttcaccccat caactcctgt 1860
gccaaccctc tcctctatgc catctttacc aaaaactttc gcagagattt cttcattctg 1920
ctgagcaagt gtggctgcta tgaaatgcaa gcccaaattt ataggacaga aacttcatcc 1980
actgtccaca acacccatcc aaggaatggc cactgtctct cagctcccag agtcaccagt 2040
ggttcactt acatacttgt ccctctaagt catttagccc aaaactaa 2088

```

&lt;210&gt; 32

&lt;211&gt; 695

&lt;212&gt; PRT

&lt;213&gt; Artificial

&lt;400&gt; 32

```

Met Ala Leu Leu Leu Val Ser Leu Leu Ala Phe Leu Ser Leu Gly Ser
1           5           10          15

```

```

Gly Cys His His Arg Ile Cys His Cys Ser Asn Arg Val Phe Leu Cys
20           25           30

```

```

Gln Glu Ser Lys Val Thr Glu Ile Pro Ser Asp Leu Pro Arg Asn Ala
35           40          45

```

```

Ile Glu Leu Arg Phe Val Leu Thr Lys Leu Arg Val Ile Gln Lys Gly
50           55          60

```

```

Ala Phe Ser Gly Phe Gly Asp Leu Glu Lys Ile Glu Ile Ser Gln Asn
65           70          75          80

```

```

Asp Val Leu Glu Val Ile Glu Ala Asp Val Phe Ser Asn Leu Pro Lys
85           90          95

```

Leu His Glu Ile Arg Ile Glu Lys Ala Asn Asn Leu Leu Tyr Ile Thr  
 100 105 110

Pro Glu Ala Phe Gln Asn Leu Pro Asn Leu Gln Tyr Leu Leu Ile Ser  
 115 120 125

Asn Thr Gly Ile Lys His Leu Pro Asp Val His Lys Ile His Ser Leu  
 130 135 140

Gln Lys Val Leu Leu Asp Ile Gln Asp Asn Ile Asn Ile His Thr Ile  
 145 150 155 160

Glu Arg Asn Ser Phe Val Gly Leu Ser Phe Glu Ser Val Ile Leu Trp  
 165 170 175

Leu Asn Lys Asn Gly Ile Gln Glu Ile His Asn Cys Ala Phe Asn Gly  
 180 185 190

Thr Gln Leu Asp Ala Val Asn Leu Ser Asp Asn Asn Asn Leu Glu Glu  
 195 200 205

Leu Pro Asn Asp Val Phe His Gly Ala Ser Gly Pro Val Ile Leu Asp  
 210 215 220

Ile Ser Arg Thr Arg Ile His Ser Leu Pro Ser Tyr Gly Leu Glu Asn  
 225 230 235 240

Leu Lys Lys Leu Arg Ala Arg Ser Thr Tyr Asn Leu Lys Lys Leu Pro  
 245 250 255

Thr Leu Glu Lys Leu Val Ala Leu Met Glu Ala Ser Leu Thr Tyr Pro  
 260 265 270

Ser His Cys Cys Ala Phe Ala Asn Trp Arg Arg Gln Ile Ser Glu Leu  
 275 280 285

His Pro Ile Cys Asn Lys Ser Ile Leu Arg Gln Glu Val Asp Tyr Met  
 290 295 300

Thr Gln Ala Arg Gly Gln Arg Ser Ser Leu Ala Glu Asp Asn Glu Ser  
 305 310 315 320

Ser Tyr Ser Arg Gly Phe Asp Met Thr Tyr Thr Glu Phe Asp Tyr Asp  
 325 330 335

Leu Cys Asn Glu Val Val Asp Val Thr Cys Ser Pro Lys Pro Asp Ala  
 340 345 350

Phe Asn Pro Cys Glu Asp Ile Met Gly Tyr Asn Ile Leu Arg Val Leu  
 355 360 365

Ile Trp Phe Ile Ser Ile Leu Ala Ile Thr Gly Asn Ile Ile Val Leu  
 370 375 380

Val Ile Leu Thr Thr Ser Gln Tyr Lys Leu Thr Val Pro Arg Phe Leu  
 385 390 395 400

Met Cys Asn Leu Ala Phe Ala Asp Leu Cys Ile Gly Ile Tyr Leu Leu  
 405 410 415

Leu Ile Ala Ser Val Asp Ile His Thr Lys Ser Gln Tyr His Asn Tyr  
 420 425 430

Ala Ile Asp Trp Gln Thr Gly Ala Gly Cys Asp Ala Ala Gly Phe Phe  
 435 440 445

Thr Val Phe Ala Ser Glu Leu Ser Val Tyr Thr Leu Thr Ala Ile Thr  
 450 455 460

Leu Glu Arg Trp His Thr Ile Thr His Ala Met Gln Leu Asp Cys Lys  
 465 470 475 480

Val Gln Leu Arg His Ala Ala Ser Val Met Val Met Gly Trp Ile Phe  
 485 490 495

Ala Phe Ala Ala Ala Leu Phe Pro Ile Phe Gly Ile Ser Ser Tyr Met  
 500 505 510

Lys Val Ser Ile Cys Leu Pro Met Asp Ile Asp Ser Pro Leu Ser Gln  
 515 520 525

Leu Tyr Val Met Ser Leu Leu Val Leu Asn Val Leu Ala Phe Val Val  
 530 535 540

Ile Cys Gly Cys Tyr Ile His Ile Tyr Leu Thr Val Arg Asn Pro Asn  
 545 550 555 560

Ile Val Ser Ser Ser Ser Asp Thr Arg Ile Ala Lys Arg Met Ala Met  
 565 570 575

Leu Ile Phe Thr Gly Phe Leu Cys Met Ala Pro Ile Ser Phe Phe Ala  
 580 585 590

Ile Ser Ala Ser Leu Lys Val Pro Leu Ile Thr Val Ser Lys Ala Lys  
 595 600 605

Ile Leu Leu Val Leu Phe His Pro Ile Asn Ser Cys Ala Asn Pro Phe  
 610 615 620

Leu Tyr Ala Ile Phe Thr Lys Asn Phe Arg Arg Asp Phe Phe Ile Leu  
 625 630 635 640

Leu Ser Lys Cys Gly Cys Tyr Glu Met Gln Ala Gln Ile Tyr Arg Thr  
 645 650 655

Glu Thr Ser Ser Thr Val His Asn Thr His Pro Arg Asn Gly His Cys  
 660 665 670

Ser Ser Ala Pro Arg Val Thr Ser Gly Ser Thr Tyr Ile Leu Val Pro  
 675 680 685

Leu Ser His Leu Ala Gln Asn  
 690 695

<210> 33

<211> 2088

<212> DNA

<213> Artificial

<400> 33

atggccctgc tctggtctc tttgctggca ttctgagct tgggctcagg atgtcatcat	60
cggatctgtc actgctctaa cagggttttt ctctgccaaag agagcaaggc gacagagatt	120
ccttctgacc tcccaggaa tgccattgaa ctgaggtttg tctcaccaa gcttcgagtc	180
atccaaaaag gtgcattttc aggatttggg gacctggaga aaatagagat ctctcagaat	240
gatgtcttgg aggtgataga ggcagatgtg ttctccaacc ttcccaaatt acatgaaatt	300
agaattgaaa aggccaacaa cctgctctac atcacccctg aggccttcca gaaccttccc	360
aaccttcaat atctgttaat atccaacaca ggtattaagc accttccaga tgttcacaag	420
attcattctc tccaaaaggc tttacttgac attcaagata acataaacat ccacacaatt	480
gaaagaaatt ctttcgtggg gctgagcttt gaaagtgtga ttctatggct gaataagaat	540

```

gggattcaag aaatacacaa ctgtgcattc aatggaaccc aactagatgc agtgaatcta    600
agcgataata ataatttaga agaattgcct aatgatgttt tccacggagc ctctggacca    660
gtcattctag atatttcaag aacaaggatc cattccctgc ctagctatgg cttagaaaat    720
cttaagaagc tgagggccag gtcgacttac aacttaaaaa agctgcctac tctggaaaag    780
cttgctgccc tcatggaagc cagcctcacc tatcccagcc attgctgtgc ctttgcaaac    840
tggagacggc aaatctctga gcttcatcca atttgcaaca aatctatttt aaggcaagaa    900
gttgattata tgactcaggc taggggtcag agatcctctc tggcagaaga caatgagtcc    960
agctacagca gaggatttga catgacgtac actgagtttg actatgactt atgcaatgaa   1020
gtgggtgacg tgacctgctc ccctaagcca gatgcattca acccatgtga agatatcatg   1080
gggtacaaca tcctcagagt cctgatatgg tttatcagca tcctggccat cactgggaac   1140
atcatagtgc tagtgatcct aactaccagc caatataaac tcacagtccc caggttcctt   1200
atgtgcaacc tggcctttgc tgatctctgc attggaatct acctgctgct cattgcatca   1260
gttgatatcc ataccaagag ccaatatcac aactatgcca ttgactggca aactggggca   1320
ggctgtgatg ctgctggctt ttctactgtc ttgccagtg agctgtcagt ctacactctg   1380
acagctatca ccttggaag atggcatacc atcacgcatg ccatgcagct ggactgcaag   1440
gtgcagctcc gccatgctgc cagtgtcatg gtgatgggct ggatttttgc ttttgagct   1500
gccctctttc ccatctttgg catcagcagc tacatgaagg tgagcatctg cctgcccag   1560
gatattgaca gccctttgtc acagctgtat gtcatgtccc tccttgtgct caatgtcctg   1620
gcctttgtgg tcatctgtgg ctgctatata cacatctacc tcacagtgcg gaacccaac   1680
atcgtgtcct cctctagtga caccaggatc gccaaagcga tggccatgct catcttcact   1740
gacttcctct gcatggcacc catttctttc ttgccattt ctgcctccct caagtgccc   1800
ctcatcactg tgtccaaagc aaagattctg ctggttctgt ttcaccccat caactcctat   1860
gccaacccct tcctctatgc catctttacc aaaaactttc gcagagattt cttcattctg   1920
ctgagcaagt gtggctgcta tgaaatgcaa gcccaaattt ataggacaga aacttcatcc   1980
actgtccaca acacccatcc aaggaatggc cactgtctct cagctcccag agtcaccagt   2040
ggttccactt acatacttgt ccctctaagt catttagccc aaaactaa                2088

```

&lt;210&gt; 34

&lt;211&gt; 695

&lt;212&gt; PRT

&lt;213&gt; Artificial

&lt;400&gt; 34

Met Ala Leu Leu Leu Val Ser Leu Leu Ala Phe Leu Ser Leu Gly Ser  
 1 5 10 15

Gly Cys His His Arg Ile Cys His Cys Ser Asn Arg Val Phe Leu Cys  
 20 25 30

Gln Glu Ser Lys Val Thr Glu Ile Pro Ser Asp Leu Pro Arg Asn Ala  
 35 40 45

Ile Glu Leu Arg Phe Val Leu Thr Lys Leu Arg Val Ile Gln Lys Gly  
 50 55 60

Ala Phe Ser Gly Phe Gly Asp Leu Glu Lys Ile Glu Ile Ser Gln Asn  
 65 70 75 80

Asp Val Leu Glu Val Ile Glu Ala Asp Val Phe Ser Asn Leu Pro Lys  
 85 90 95

Leu His Glu Ile Arg Ile Glu Lys Ala Asn Asn Leu Leu Tyr Ile Thr  
 100 105 110

Pro Glu Ala Phe Gln Asn Leu Pro Asn Leu Gln Tyr Leu Leu Ile Ser  
 115 120 125

Asn Thr Gly Ile Lys His Leu Pro Asp Val His Lys Ile His Ser Leu  
 130 135 140

Gln Lys Val Leu Leu Asp Ile Gln Asp Asn Ile Asn Ile His Thr Ile  
 145 150 155 160

Glu Arg Asn Ser Phe Val Gly Leu Ser Phe Glu Ser Val Ile Leu Trp  
 165 170 175

Leu Asn Lys Asn Gly Ile Gln Glu Ile His Asn Cys Ala Phe Asn Gly  
 180 185 190

Thr Gln Leu Asp Ala Val Asn Leu Ser Asp Asn Asn Asn Leu Glu Glu  
 195 200 205

Leu Pro Asn Asp Val Phe His Gly Ala Ser Gly Pro Val Ile Leu Asp  
 210 215 220

Ile Ser Arg Thr Arg Ile His Ser Leu Pro Ser Tyr Gly Leu Glu Asn

225		230		235		240
Leu Lys Lys Leu Arg Ala Arg Ser Thr Tyr Asn Leu Lys Lys Leu Pro		245		250		255
Thr Leu Glu Lys Leu Val Ala Leu Met Glu Ala Ser Leu Thr Tyr Pro		260		265		270
Ser His Cys Cys Ala Phe Ala Asn Trp Arg Arg Gln Ile Ser Glu Leu		275		280		285
His Pro Ile Cys Asn Lys Ser Ile Leu Arg Gln Glu Val Asp Tyr Met		290		295		300
Thr Gln Ala Arg Gly Gln Arg Ser Ser Leu Ala Glu Asp Asn Glu Ser		305		310		315
Ser Tyr Ser Arg Gly Phe Asp Met Thr Tyr Thr Glu Phe Asp Tyr Asp		325		330		335
Leu Cys Asn Glu Val Val Asp Val Thr Cys Ser Pro Lys Pro Asp Ala		340		345		350
Phe Asn Pro Cys Glu Asp Ile Met Gly Tyr Asn Ile Leu Arg Val Leu		355		360		365
Ile Trp Phe Ile Ser Ile Leu Ala Ile Thr Gly Asn Ile Ile Val Leu		370		375		380
Val Ile Leu Thr Thr Ser Gln Tyr Lys Leu Thr Val Pro Arg Phe Leu		385		390		395
Met Cys Asn Leu Ala Phe Ala Asp Leu Cys Ile Gly Ile Tyr Leu Leu		405		410		415
Leu Ile Ala Ser Val Asp Ile His Thr Lys Ser Gln Tyr His Asn Tyr		420		425		430
Ala Ile Asp Trp Gln Thr Gly Ala Gly Cys Asp Ala Ala Gly Phe Phe		435		440		445
Thr Val Phe Ala Ser Glu Leu Ser Val Tyr Thr Leu Thr Ala Ile Thr		450		455		460
Leu Glu Arg Trp His Thr Ile Thr His Ala Met Gln Leu Asp Cys Lys		465		470		475
						480

Val Gln Leu Arg His Ala Ala Ser Val Met Val Met Gly Trp Ile Phe  
 485 490 495

Ala Phe Ala Ala Ala Leu Phe Pro Ile Phe Gly Ile Ser Ser Tyr Met  
 500 505 510

Lys Val Ser Ile Cys Leu Pro Met Asp Ile Asp Ser Pro Leu Ser Gln  
 515 520 525

Leu Tyr Val Met Ser Leu Leu Val Leu Asn Val Leu Ala Phe Val Val  
 530 535 540

Ile Cys Gly Cys Tyr Ile His Ile Tyr Leu Thr Val Arg Asn Pro Asn  
 545 550 555 560

Ile Val Ser Ser Ser Ser Asp Thr Arg Ile Ala Lys Arg Met Ala Met  
 565 570 575

Leu Ile Phe Thr Asp Phe Leu Cys Met Ala Pro Ile Ser Phe Phe Ala  
 580 585 590

Ile Ser Ala Ser Leu Lys Val Pro Leu Ile Thr Val Ser Lys Ala Lys  
 595 600 605

Ile Leu Leu Val Leu Phe His Pro Ile Asn Ser Tyr Ala Asn Pro Phe  
 610 615 620

Leu Tyr Ala Ile Phe Thr Lys Asn Phe Arg Arg Asp Phe Phe Ile Leu  
 625 630 635 640

Leu Ser Lys Cys Gly Cys Tyr Glu Met Gln Ala Gln Ile Tyr Arg Thr  
 645 650 655

Glu Thr Ser Ser Thr Val His Asn Thr His Pro Arg Asn Gly His Cys  
 660 665 670

Ser Ser Ala Pro Arg Val Thr Ser Gly Ser Thr Tyr Ile Leu Val Pro  
 675 680 685

Leu Ser His Leu Ala Gln Asn  
 690 695

<210> 35

<211> 36

<212> DNA

<213> Artificial

<400> 35

gatcaagctt ccatggcgtg ctgcctgagc gaggag

36

<210> 36

<211> 53

<212> DNA

<213> Artificial

<400> 36

gatcggatcc ttagaacagg ccgcagtcct tcagggtcag ctgcaggatg gtg

53

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**